This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Symptom profile of severe mental illness and adverse health outcomes

Jayatilleke, Nishamali Lakshani

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

Symptom profile of severe mental illness and adverse health outcomes

Nishamali Jayatilleke

Doctor of Philosophy in Psychological Medicine

Department of Psychological Medicine,

Institute of Psychiatry, Psychology and Neuroscience,

King's College London

June 2017

Abstract

Background

Individuals with severe mental illness (SMI- schizophrenia, schizoaffective disorder and bipolar disorder) experience higher levels of morbidity and mortality than the general population. An important policy goal is to reduce this gap. Investigating the contributory role of physical illness is of key importance to unpacking the associations between SMI and detrimental outcomes, such as premature mortality and frequent hospital admissions. This thesis builds on recent advances to data extraction technologies to investigate the above.

Objectives

- To describe the relative contributions of major disease groups to the gap in life expectancy between individuals with SMI and the general population.
- To describe the most common reasons for admission to non-psychiatric hospitals by individuals with SMI and the relative frequencies of these admissions compared to the general population.
- As a proof of principle for ascertaining meaningful symptom profiles from routine mental health record text fields, to describe the prospective association between number of recorded negative symptoms and mental healthcare outcomes (admission, duration of admission, and readmission) among individuals with schizophrenia.

- To describe the association with mortality and hospitalisation for each of six symptom dimensions (positive, negative, manic, disorganisation, catatonic and depressive) extracted from the clinical records of individuals with SMI.

Methods

Information for SMI cohorts were derived from the Clinical Record Interactive Search (CRIS), a de-identified electronic patient records data resource. Data on mortality were extracted using existing linkages between CRIS and death certification (Office for National Statistics). Life expectancy estimates were used to explain the contributions of specific causes of death to the gap. Using Hospital Episode Statistics data, frequencies of and causes for non-psychiatric hospital admissions in SMI were compared to those in the catchment general population. Symptoms within clinical record text fields were ascertained using a range of natural language processing algorithms, and were assessed for their associations with mortality and hospitalisation outcomes.

Results

Natural causes accounted for 79.2% of lost life-years in women and 78.6% in men. Deaths from circulatory disorders accounted for more life-years lost in women than men (22.0% versus 17.4%, respectively), as did deaths from cancer (8.1% versus 0%), but the contribution from respiratory disorders was lower in women than men (13.7% versus 16.5%).

Commonest discharge diagnosis categories were urinary conditions, digestive conditions, unclassified symptoms, neoplasms, and respiratory conditions. SARs were raised for most major categories, except neoplasms where risk was significantly lower. Hospitalisation risks were specifically higher for poisoning and external causes, injury, endocrine/metabolic conditions, haematological, neurological, dermatological, infectious and non- specific ('Z-code') causes. The five commonest specific ICD-10 diagnoses at discharge were 'chronic renal failure' (N18), a non-specific code (Z04), 'dental caries' (K02), 'other disorders of the urinary system' (N39), and 'pain in throat and chest' (R07), all of which were higher than expected (SARs ranging 1.57–6.66).

Proof of concept analyses showed negative symptoms were associated with younger age, male gender and single marital status, and with increased likelihood of hospital admission (OR 1.24, 95% CI 1.10 to 1.39), longer duration of admission (β-coefficient 20.5 days, 95% CI 7.6–33.5), and increased likelihood of readmission following discharge (OR 1.58, 95% CI 1.28 to 1.95).

Cox regression analyses detected significant effect of positive (HR 1.08, 95% CI 1.03- 1.16), negative (HR 1.09 95% CI 1.02- 1.16) and catatonic (HR 1.09 95% CI 1.03- 1.16) symptoms on mortality with adjustment of age, sex, employment, marital status and ethnicity. Linear regression analyses detected significant effect of manic (β-coefficient 0.09, 95% CI 0.02- 0.15), catatonic (β-coefficient 0.08, 95% CI 0.02- 0.15) and depressive (β-coefficient 0.14,95% CI 0.08- 0.21) symptoms on admission to non-mental health hospitals.

Implications

Clinically, findings from this thesis confirm that SMI has a substantial negative impact on physical health, associated with increased mortality and morbidity. Of clinical relevance, the findings showed differences in burden depending on types of symptoms recorded in health records. From a policy context, the gap in life expectancy and increased non-psychiatric hospitalisation, accounted for by a broad range of causes, need to be addressed systematically. Interventions should focus on a whole system approach to improve health benefits for individuals of SMI.

Acknowledgements

I am indebted to my supervisors, Professor Robert Stewart and Dr Chin-Kuo Chang for their mentorship throughout this thesis. I would also like to acknowledge the support from BRC Nucleus Team including my clinical informatician colleagues and other academics.

The work described in this thesis used the Clinical Records Interactive Search (CRIS) system which was funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, with a joint infrastructure grant provided by the Guy's and St. Thomas' Charity and the Maudsley Charity (grant number: BRC-2011-10035).

The research was carried out with ethical approval. CRIS was developed to allow search and retrieval of anonymised but complete medical records within a robust patient-led governance framework (Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, Fernandes A, Hayes RD, Henderson M, Jackson R, Jewell A, Kadra G, Little R, Pritchard M, Shetty H, Tulloch A 2016; Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang CK, Jackson RG, Roberts A, Tsang J, Soncul M, Liebscher J, Stewart R 2013) and is approved as a dataset for secondary analysis (Oxfordshire Research Ethics Committee C, reference 08/H0606/71+5).

Contents

List of tables

List of figures

Glossary of abbreviations

BRC	Biomedical Research Centre
CBT	Cognitive behavioural therapy
CI	Confidence interval
CRIS	Clinical Record Interactive Search
DSM	Diagnostic and Statistical Manual of Mental Disorders
GATE	General Architecture for Text Engineering
HES	Hospital Episode Statistics
HoNOS	Health of the Nation Outcome Scales
HR	Hazard ratio
ICD	International Classification of Diseases
NHS	UK National Health Service
NICE	UK National Institute of Health and Care Excellence
NIHR	National Institute of Health Research
NLP	Natural Language Processing
ONS	Office for National Statistics
OR	Odds ratio
PANSS	Positive and Negative Syndrome Scale
PJS	Patient Journey System
RR	Risk ratio
SAR	Standardised admissions ratio
SLaM	South London and Maudsley NHS Foundation Trust
SMI	Severe mental illness
SMR	Standardised morbidity ratio
SQL	Structured query language
UK	United Kingdom
WHO	World Health Organisation

Outline of the thesis

Chapter 1- This chapter aims to introduce the conditions that collectively form severe mental illness, its domains of psychopathology and dimensions of psychosis. Electronic health records (EHRs) are recognised as a valuable source of data and the use of natural language processing as a methodology to extract meaningful approaches to obtain clinical information. This can be viewed as an opportunity in the context of the described implications of physical illness and inequalities that individuals with severe mental illness experience. The chapter concludes with a summary of key issues identified.

Chapter 2- Provides a literature review of predictors of mortality for individuals with serious mental illnesses.

Chapter 3- Provides a summary of the gaps in the literature identified.

Chapter 4- Presents overall thesis objectives and individual project rationale and aims.

Chapter 5- Describes the main methods used in the thesis and planned analyses.

Chapter 6- Investigates the major disease groups that contribute to the gap in life expectancy between individuals with severe mental illness and the general population. Given the life expectancy gap between people with severe mental illness and the general population persists and may even be widening, this chapter aims to estimate contributions of specific causes of death to the gap.

The chapter is based on the following peer-reviewed paper-

Jayatilleke N, Hayes RD, Dutta R, Shetty H, Hotopf M, Chang CK, Stewart R. Contributions of specific causes of death to lost life expectancy in severe mental illness. European Psychiatry. 2017 Mar 17;43:109-115.

Chapter 7- Investigates the commonly recorded reasons for acute care hospitalisation in people with severe mental illness. Given that severe mental illness is associated with worse general health, but reasons for admissions to general hospitals have received little investigation, this chapter aims to compare frequencies of non-psychiatric hospital admission by major diagnostic categories between people with SMI and the general population.

The paper reporting this chapter is currently under peer review

Jayatilleke N, Hayes RD, Chang CK, Stewart R. Acute general hospital admissions in people with severe mental illness. [In submission].

Chapter 8- Investigates the relationship between negative symptoms identified in the clinical records of a large sample of patients with schizophrenia using natural language processing and clinical outcomes.

The chapter is based on the following peer-reviewed paper-

Patel R*, Jayatilleke N*, Broadbent M, Chang CK, Foskett N, Gorrell G, Hayes RD, Jackson R, Johnston C, Shetty H, Roberts A, McGuire P, Stewart R. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. BMJ Open. 2015 Sep 7;5(9):e007619.

* contributed equally

Chapter 9- Investigates symptoms in the electronic clinical records of a large mental healthcare cohort of patients with schizophrenia, schizoaffective disorder, or bipolar disorder using natural language processing and assesses their relationship with all-cause mortality and hospitalisation index.

The paper reporting this chapter is currently under peer review

Jayatilleke N, Chang CK, Stewart R. Symptoms identified in electronic medical records for individuals with severe mental illness by natural language processing in relation to general mortality and hospitalisation [In submission].

Chapter 10- Considers all findings from the previous chapters in relation to existing research. I also discuss the strengths and limitations of the thesis and consider how my

findings fit into the wider context of patients with serious mental illnesses. I further discuss other possible implications and ideas for further research.

1 Introduction

1.1 Chapter outline

This chapter aims to introduce the conditions that collectively form severe mental illness, its domains of psychopathology and dimensions of psychosis. All mental disorders can be debilitating; however, for the purposes of this thesis, severe mental illness (SMI) refers to schizophrenia, schizoaffective disorder and bipolar disorders. Despite psychiatric diagnoses reportedly having a high reliability (Van Os et al, 2003), there is substantial heterogeneity within SMI diagnoses and the validity of the constructs has been frequently questioned. Psychotic symptoms are to some extent unique to individuals and there are wide individual differences in the severity and course of the underlying disorders and level of functioning in individuals with the same disorder (Cowen 2018). To answer my study questions, I have extracted the data from a large electronic health record (EHR) based case register in London and the context for this will also be introduced in this chapter. EHRs are a relatively novel source of data whose value and application are just beginning to be appreciated (Stewart R, et al 2009; Fernandes AC, et al 2013; Perera G, et al 2016). The chapter concludes with a summary of key issues identified.

1.2 Severe mental illness and its constituent disorders

1.2.1 Component ‘SMI’ conditions

The International Classification of Diseases, 10th edition (ICD-10) diagnoses of schizophrenia (F20.x), schizoaffective disorder (F25.x) and/or bipolar disorder (F31.x) that collectively form SMI have substantial impacts on individuals’ cognitive, affective, behavioural and physical state (Stahl 2013) with the potential to cause long-term disability.

Schizophrenia is characterised by profound disruptions in thinking, affecting language, perception, and sense of self (Cowen 2018). Its manifestations include psychotic experiences, such as auditory hallucinations and delusions, but is often also accompanied by so-called ‘negative symptoms’ and a broader impairment in functioning resulting in a reduced capability to earn a livelihood, the disruption of studies and social isolation (WHO 2017).

Bipolar disorder is characterised by episodes of mania (abnormally elevated mood or irritability and related symptoms with severe functional impairment or psychotic symptoms for 7 days or more) or hypomania (abnormally elevated mood or irritability and related symptoms with decreased or increased function for 4 days or more) and episodes of depressed mood (Cowen 2018, NICE 2014).

Schizoaffective disorder is an episodic disorder characterised by both affective and schizophrenic symptoms prominent within the same episode of illness: mostly simultaneously, but at least within a few days of each other (WHO 2017).

These constituent disorders and associated illness episodes can be further sub-categorised, as displayed in Table 1.1 for the ICD-10 classification system.

Table 1:1 - Schizophrenia, schizoaffective disorder and bipolar disorder classification

International Classification of Diseases version 10 (ICD-10)
F20 Schizophrenia F20.0 Paranoid schizophrenia F20.1 Hebephrenic schizophrenia F20.2 Catatonic schizophrenia F20.3 Undifferentiated schizophrenia F20.4 Post-schizophrenic depression F20.5 Residual schizophrenia F20.6 Simple schizophrenia F20.8 Other schizophrenia F20.9 Schizophrenia, unspecified
F25 Schizoaffective disorders F25.0 Schizoaffective disorder, manic type F25.1 Schizoaffective disorder, depressive type F25.2 Schizoaffective disorder, mixed type F25.8 Other schizoaffective disorders F25.9 Schizoaffective disorder, unspecified
F31 Bipolar affective disorder F31.0 Bipolar affective disorder, current episode hypomanic F31.1 Bipolar affective disorder, current episode manic without psychotic symptoms F31.2 Bipolar affective disorder, current episode manic with psychotic symptoms (.20 With mood-congruent psychotic symptoms .21 With mood-incongruent psychotic symptoms)

F31.3 Bipolar affective disorder, current episode mild or moderate depression (.30 Without somatic syndrome .31 With somatic syndrome)

F31.4 Bipolar affective disorder, current episode severe depression without psychotic symptoms

F31.5 Bipolar affective disorder, current episode severe depression with psychotic symptoms

(.50 With mood-congruent psychotic symptoms .51 With mood-incongruent psychotic symptoms)

F31.6 Bipolar affective disorder, current episode mixed

F31.7 Bipolar affective disorder, currently in remission

F31.8 Other bipolar affective disorders

F31.9 Bipolar affective disorder, unspecified

1.2.2 The aetiology of SMI

The exact aetiology for this group of conditions is not clear but is considered to be multifactorial (Cowen 2018). Biological theories have suggested a combination of genetic factors, structural and functional abnormalities in the brain and environmental contributions (Bellmaker R 2004; Picchioni & Murray 2007; Van Os & Kapur 2009).

Family, twin and adoption studies indicate an important genetic contribution; however, it appears that a 'liability to develop' the disorder is what is inherited rather than a certainty of getting it, since even monozygotic twins show concordance rates substantially less than 100% (Cowen 2018). Epidemiological studies have confirmed that psychiatric disorders more generally tend to co-occur within families (Sebat J et al. 2009). Thus, if a person has a first-degree relative with schizophrenia, while their risk of schizophrenia is ten times higher than the average person's, their risk of bipolar disorder or autism or epilepsy is also dramatically increased (Lichtenstein et al. 2009). If these disorders are caused by single mutations then the implication is that any such mutation may result in a spectrum of symptoms in one individual, which leads to a diagnosis of disorder A, and in a different spectrum in another individual, which leads to a diagnosis of disorder B. In all these cases, such mutations are associated with increased risk not to just one disorder but to many (Sebat J et al. 2009).

There may be lowered functioning in the prefrontal areas of the brain and enlarged ventricles suggesting localised under-development(Cowen 2018). There is also some

evidence for the involvement of neurotransmitters, and alteration of neurotransmitter levels remains the mainstay of treatment; however, it is unlikely that a dysfunction of any single neurotransmitter system can account for a disorder as complex as schizophrenia, or that the pathologies identified to date only affect one neurotransmitter system. It has been suggested that schizophrenia is a "disconnection" syndrome which affects the organisation of neurones in the brain (Van Os & Kapur 2009). Clinically, all three conditions are heterogeneous and this may point to heterogeneous aetiology (Cowen 2018). It seems that genetics, neurodevelopmental problems, neurochemistry and abnormal connectivity, as well as psychosocial stressors (van Os et al. 2010), probably all contribute to developing the typical clinical presentation. Environmental contributions have an important role. In particular, for bipolar disorder, recent life events and interpersonal relationships contribute to the likelihood of onset and recurrence of bipolar mood episodes (Alloy et al. 2005). There is also evidence that high frequency of early life stressors are contributory (Leverich & Post 2006).

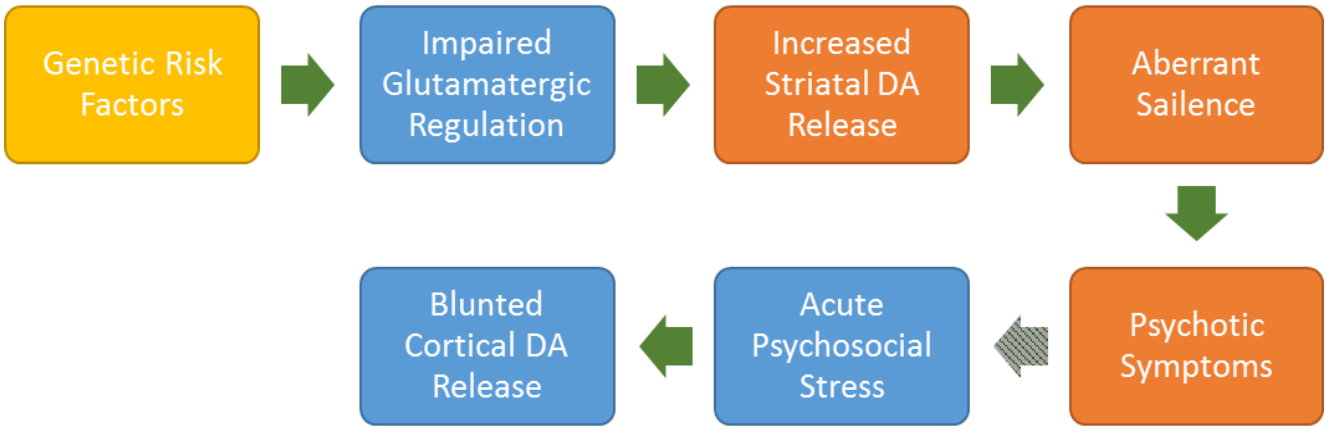
Several neurotransmitter system genes have been examined for their role in brain functions and their reported genetic associations with cognition (Cowen 2018). As a result, several genetic variations are emerging as having potential effects on neurocognitive deficits (Zai et al. 2017); these include genes with roles in the neurotrophic, serotonin, cell adhesion, and sodium channel systems. Another incomplete explanation is the dopamine hypothesis which is founded on two sources of evidence. First, it is recognised that dopaminergic agonists and stimulants can induce psychosis in healthy individuals and can worsen psychosis in an individual with SMI (Howes 2009). Second, most effective antipsychotic drugs have

recognised effects on the dopamine system. Furthermore, while dopaminergic functioning is already dysregulated in prodromal individuals who later progress to schizophrenia, it is not as marked as in individuals with the disorder, and there is further dysregulation from the prodrome to psychosis(Howes et al. 2016). It remains to be determined whether there is enrichment of genetic signal in other restricted gene sets relating to dopaminergic function, such as signal transduction and postsynaptic signalling. It is also possible that dopaminergic genes play a more prominent role in clinical subsets or in cases defined based on drug response(Howes et al. 2017). A third addition to this hypothesis accounts for the multiple environmental and genetic risk factors for the condition and proposes that these interact to funnel through one final common pathway of excess dopaminergic activity (Howes 2009). Figure 1.1 shows a model of genetic factors leading to psychotic symptoms through dopaminergic dysregulation. Genetic factors that lead to impaired regulation which leads to increased dopamine release. In addition how the individual responds to acute stress, with less blunted cortical dopamine and increased striatal dopamine leads to further sensitisation of the dopamine system(Howes et al. 2017).

The combination of dopamine-related disturbances with other environmental and genetic factors may provide a framework linking the neurochemistry to symptoms experienced by individuals. It may also reveal why many disparate risk factors and functional and structural abnormalities are associated, for example, with schizophrenia but are not specific to schizophrenia, as mentioned above.

Figure 1-1 -Model integrating genetic and environmental factors, dopaminergic dysregulation, and the development of psychotic symptoms

(Abstracted from (Howes et al. 2017))



Note- DA denotes dopamine

1.2.3 Epidemiology & treatment of SMI

SMI is relatively rare but for those affected the impact can be lifelong. Schizophrenia affects around 0.3–0.7% of people at some point in their life while the lifetime prevalence of bipolar I disorder (mania and depression) is estimated at 1% of the adult population, and bipolar II disorder (hypomania and depression) affects approximately 0.4% of adults. Schizoaffective disorder is estimated to occur in 0.5 to 0.8% of people at some point in their life (Van Os & Kapur 2009; Ketter 2010).

SMI disorders are usually treated with an individually tailored combination of therapy and medication. Early and prompt intervention is recommended for all disorders. In the UK National Health Service, most people with SMI are treated by community mental health teams with an aim to provide day-to-day support and treatment while enabling independence for the individual (Cowan 2018). A combination of professionals including mental health nurses, mental health occupational therapists, social workers, counsellors and psychotherapists, psychologists and psychiatrists may be involved. Treatment in the UK is generally encompassed within a 'care programme approach'. An assessment of needs, care plan creation, nominated key worker and regular reviews are all fundamental to this model. People who have serious psychotic symptoms as the result of an acute episode may require a more intensive level of care including hospitalisation or care from crisis resolution teams. Compulsory detention may be required for the most serious acute episodes under the Mental Health Act (2007) if justified in the interests of the person's own health and safety or to protect others.

For schizophrenia, antipsychotic drugs are usually recommended as the initial treatment for the symptoms of an acute episode, and the mechanism of action generally includes blocking the effect of the neurotransmitter dopamine. Antipsychotics are usually effective on feelings of anxiety or aggression within a few hours of use, but may take several days or weeks to reduce other symptoms, such as hallucinations or delusional thoughts. Psychological treatment can help cope with the symptoms of hallucinations or delusions better. They can also help treat some of the negative and depressive symptoms, such as apathy or anhedonia. Psychological treatments for schizophrenia in conjunction with antipsychotic medication is recommended treatment approach (NICE 2014), and include cognitive behavioural therapy (CBT), family therapy and arts therapy. CBT for psychosis aims to help the patient identify thinking patterns that are causing unwanted feelings and behaviour, and learn to replace this thinking with more realistic and useful thoughts. Family therapy is a way of helping the family cope better with the condition involving a series of informal meetings discussing information about schizophrenia, exploring ways of supporting somebody with schizophrenia and deciding how to solve practical problems that can be caused by the symptoms of schizophrenia. Arts therapies are designed to promote creative expression to help develop socialisation, and have been shown to alleviate negative symptoms of schizophrenia in some people (NHS Choices 2017; Cowen 2018; NICE 2009). However, the evidence for adjunctive treatment modalities still remain questionable as referring individuals with the condition does not lead to improved global functioning(Crawford et al. 2012).

For bipolar disorders, drug treatment includes mood stabilizers such as lithium, sodium valproate, lamotrigine, and carbamazepine, and second-generation antipsychotic agents

(Ketter 2010) with potential co-administration of antidepressants, anxiolytics/hypnotics, and other anticonvulsants. Antipsychotic agents and mood stabilisers provide the primary means for managing an acute manic or hypomanic relapse, with continuation of mood stabilisers with or without an antipsychotic agent to prevent relapse following recovery. Depressive episodes may be managed with antidepressant treatment, although with close monitoring for mania/hypomania consequently. Psychotherapeutic interventions may be used in the management of depressive episodes and to aid recovery and prevent relapse during periods of remission.

For schizoaffective disorder, treatment can be a combination of that used for schizophrenia and bipolar disorder, however, it needs to be tailored to the individual's specific symptoms and situation (Cowen 2018). Drug treatments, along with more psychosocial therapies, are often necessary to successfully treat schizoaffective disorder. Drug therapy usually can stop the patient's psychosis, but often only social and occupational rehabilitation therapies can overcome any associated problems such as unemployment, poverty and homelessness (Cowen 2018).

1.3 Domains of psychopathology and dimensions of psychosis

1.3.1 Phenomenology of psychosis

The construct of psychosis has long been debated. Several theories have been put forward dating back to the early 20th century. The multiple avenues led by the debate illustrates that the scientific status was never firmly established, anticipating continuing requests to rethink SMI. Kraeplin was the first to dichotomise psychotic disorders into what were termed 'manic depressive insanity' and 'dementia praecox' (Kraeplin 1893). This distinction was primarily based on the course of the disorders where manic depression was said to have more of a relapsing and remitting course, with returns to health and full functioning between episodes, whereas dementia praecox was proposed as characterised by a more chronic and declining picture. The term 'dementia' implied chronicity and praecox meant 'of the young'. The term 'schizophrenia' was later introduced by Bleuler to replace dementia praecox, partly due to the fact that it did not only occur in younger people (Bleuler 1963). Bleuler continued to consider its course as a physical disease process but established that not everybody with the disorder underwent a progressive decline. Bleuler specified its defining symptoms as affective flattening, looseness of associations between mental functions, ambivalence and autism – i.e. using the more negative psychotic symptoms as defining features; he regarded hallucinations and delusions as secondary symptoms (Heckers 2011). The term "bipolar" — which signifies the polar opposites of mania and depression—first appeared in the American Psychiatric Association's (AMA) Diagnostic and Statistical Manual of Mental Disorders (DSM) in its third revision in 1980 (Ketter TA 2010).

Jaspers' model of “layers” of mental disorders is also of relevance, in which it was proposed that neurotic, endogenous, and organic mental disorders reflected three different layers of mental disorders, and psychotic symptoms could be found on both endogenous as well as organic levels (Jaspers 1973). The loss of reality which underpins hallucinations and delusions became important, and the term “psychosis” has been used variably to denote a core syndrome of hallucinations, delusions, and disordered thinking, or in a wider sense, to encompass all severe mental disorders. On this background of such clinical diversity and variability, Schneider introduced a ranking of psychotic symptomatology, bringing into the discussion the notion that when diagnosing and classifying mental disorders, some psychotic symptoms may be more important than others (Schneider 1959). Table 1.2 lists Schneider’s first rank symptoms.

Table 1:2 - Schneider's first rank symptoms

Schneider's first rank symptoms (1959)

Auditory hallucinations:

1. Audible thoughts: hearing one's thoughts spoken aloud.

2. Voices arguing: hearing two or more voices in disagreement or discussion. usually about the person referred to in the third person.

3. Voices commenting on one's actions as they occur.

Ego boundary disturbance/ other perceptual and delusional experiences:

4. Somatic passivity: the experience of bodily sensations imposed by an external agency – not simply the delusional interpretation of a bodily sensation.

5. Thought withdrawal: experience of thoughts being taken from the mind by an external force.

6. Thought insertion: that thoughts have literally been inserted into one's mind by an external agency.

7. Thought broadcasting: the experience of thoughts not being contained within the mind – that they are literally outside of the self in the public domain.

8. Made feelings: the experience of feelings that have been literally imposed upon the person.

9. Made impulses: an impulse to carry out an action that is felt not to belong to the person (although the actual performance of the act is).

10. Made volitional acts: experience of one's actions being controlled by an external force.

11. Delusional perception.

In practice, the difficulty has been that many of the psychotic symptoms in the different disorders overlap, and diagnosis based on the more negative symptoms has not led to reliable discernment between illness and normality. Kraepelin's initial distinctions were based on observed course of illness, Bleuler's approach focused on negative symptoms to distinguish schizophrenia, while Schneider's alternative focus on more positive symptoms has become widely used as a set of experiences that are strongly indicative of schizophrenia (Gaebel & Zielasek 2015). However, Hoch in 1912 and Kirby in 1913, both described patients with mixed features of schizophrenia and affective disorders which later established as schizoaffective disorder as their patients did not have the deteriorating course of dementia praecox but more like Kraeplin's manic depressive psychosis group (Yogeswary K 2014). Therefore, the questionable extent to which the three disorders could be distinguished in all cases was recognised from an early stage.

1.3.2 The overlap between psychotic disorders

Schizophrenia and bipolar disorder share many individual symptoms with certain combinations of these symptoms co-occurring often (Tondo et al. 2016). The distinction between these two conditions may be clinically useful for treatment and prognosis. However, they can be considered as being on two ends at either end of a spectrum, with the intervening spaces filled in with a variety of other conditions or diagnostic labels including schizoaffective disorder. While the ends of the spectrum might be quite different, each move along it might be more dimensional than categorical. Recent genetic findings are much more congruent with a dimensional distinction than a categorical one and suggest substantial shared aetiology between psychotic disorders. Insel (Insel et al. 2010) suggests that within each diagnostic category there is a considerable degree of heterogeneity and that the diagnostic category in itself provides little information about future clinical outcomes. Van Os (Van Os et al. 1997) suggests that there are overlapping genetic, neurobiological and clinical features between different categories of mental disorder. The genetic commonality has been further established as specific single-nucleotide polymorphisms are associated with a range of psychiatric disorders of childhood onset or adult onset which include bipolar disorder and schizophrenia (Cross-Disorder Group of the Psychiatric Genomics Consortium. 2013). Schizophrenia, in particular, is a highly heritable disorder with genetic risk is conferred by a large number of alleles. Also independent of genes expressed in brain, genes expressed in tissues are also associated which have immunity functions, providing support for a speculated link between the immune system and schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014).

1.2.2 Symptom classification, factor analysis approaches and dimensions of SMI

In mental health research and clinical practice, it is often argued that the symptoms expressed by a patient in the course of their illness represent a more useful description of the disorder and indications for intervention than the concept of a diagnosis (Adam 2013; Chmielewski et al. 2014). While common conditions in mental health are represented in classification taxonomies such as the International Classification of Diseases (ICD) and Diagnostic and Statistical Manual (DSM) systems, the symptomatology of a condition is more often used in clinical practice to determine an appropriate treatment plan. This is due to the broad symptomatic manifestations of mental disorders, in the sense that, at a given time, a patient assigned a diagnosis (such as schizophrenia) can present with all, many or very few of the symptoms associated with the condition. This is particularly pertinent to clinical practice where diagnoses are not necessarily assigned using research criteria. The problems of diagnostic semantics are especially apparent in SMI where controversy is compounded by the high frequency of mental health co-morbidities and shortcomings in our current understanding of the biological underpinnings of mental disorders which in turn limit our ability to sub-classify the conditions.

The lack of genetic and other objective tests for many mental disorders has led to a requirement for detailed, interpersonal observation of patients, cumulating in pragmatic symptomatology-based assessments (Anon 1983; Kirkpatrick B, Strauss GP 2011; Kring 2013; Kay et al. 1987; Axelrod et al. 1993). A common task in health research is to group patients with similar conditions into appropriate cohorts, which will almost inevitably require

ascertaining common factors pertinent to their disorder (Hersh 2007; Embi et al. 2005). The case for identifying symptoms of SMI as a source of data for mental health research is driven by widely recognised deficiencies of diagnostic categories alone for capturing mental disorders or providing adequate classes with which to cluster groups of patients for research or intervention. This is compounded by the lack of an inventory to capture symptomatology, as most research instruments would be considered overly cumbersome for routine clinical application outside specialist services. Furthermore, even if a fully structured instrument was identified as acceptable for use in initial assessment, obtaining real-time repeated measurements would present even more substantial challenges.

The composition of psychosis from varying numbers of symptoms has led to the suggestion of deconstructing the entity according to these components. Factor analysis has been the most common approach in this respect: a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors (Yong 2013). Factor analysis aims to deal with data sets where there are large numbers of observed variables that are thought to reflect a smaller number of underlying/latent variables and varying solutions have been proposed for psychosis. There have been several factor analyses published. The Positive and Negative Syndrome Scale (PANSS) is frequently referenced and is a 30-item rating scale designed to assess dimensions of schizophrenia symptoms (Kay et al. 1987). Its items were originally grouped into scales for Positive Symptoms (7 items), Negative Symptoms (7 items), and General Psychopathology (16 items). Several factor-analytic studies have suggested that a five-factor model better captures PANSS structure in schizophrenia samples. In these studies, smaller

groupings of items represent Positive and Negative symptoms, with the remainder grouped to either Disorganised/Concrete, Excited or Depressed which over time does not appear to have shifted hugely from initial descriptions (Bell MD et al 1994; Wallwork et al. 2012). If psychotic symptoms in the general population are taken into account, depressive and manic symptoms also come into play, reflecting the occurrence of signs and symptoms of psychosis in mental disorders (Gaebel & Zielasek 2015). A recent review of available studies on the dimensional structure of psychosis concluded that there is relatively consistent evidence, with the majority of the studies finding four or five dimensions, and with positive, negative, disorganisation, and affective symptom dimensions most frequently reported (Potuzak et al. 2012). Russo (2014) also concluded that symptoms of psychosis assume a multidimensional hierarchical structure which was stable over time and which showed good concordance with categorical diagnoses. The combined use of dimensional and categorical approaches to psychotic disorders was therefore concluded to be of clinical and research utility (Russo et al. 2014).

A cluster of clinical symptoms encompassing, in several possible compositions, the psychopathologic domains of delusions, hallucinations, and disorganised thinking, supplemented by affective domains, is therefore the core of psychosis. These symptoms occur in different mental disorders as currently characterised with considerable overlap, although there are clearly symptoms that occur more often in schizophrenia compared with affective psychoses. This may indicate that the causes and pathological mechanisms of psychotic symptoms in affective disorders are different from those in schizophrenia and related disorders. These findings obviously raise some doubts about the clinical validity of

traditional diagnostic categories. On the one hand, diagnoses may be arbitrarily defined constructs of limited usefulness that actually hamper medical practice and research. Alternatively, they may describe truly distinct phenotypic end-points reflecting real differences in current pathophysiology, despite similar mechanistic starting points.

As described above, dimensional assessments capture potentially more meaningful variation in the combinations and severity of symptoms, which may help with treatment planning and the prediction of course and outcome (Allardyce, J. et al. 2007). Studies investigating associations of categorical and dimensional models with different clinical parameters have reported the superiority of dimensions over diagnostic equivalents to predict clinical course, outcome and treatment response (Van Os et al. 1997). The relative clinical utility of symptom dimensions versus conventional diagnostic categories in individuals with SMI with symptom dimensions rated at presentation better predicting outcomes such as time to first remission than a categorical diagnosis (Ajnakina et al. 2018). DSM-5 also takes a more dimensional approach, which proposes that a patient who presents with the signs and symptoms of psychosis should be assessed along eight dimensions: the five domains that define schizophrenia spectrum disorder, hallucinations, delusions, disorganised thought (speech), disorganised or abnormal motor behaviour (including catatonia) and negative symptoms as well as cognition, depression, and mania (American Psychiatric Association 2013). Each DSM-5 dimension should be assessed on a five-point scale ranging from 0 which denotes not present to 4 which denotes present and severe. At present, though, it is not clear how these dimensions will map onto models of human behaviour such as decision making and reward behaviour or neural circuits

(Gigerenzer 2011; Der-Avakian 2012). However, this initial proposal to integrate dimensions into clinical practice can set the stage for future alignment of health care as well as research.

1.4 Physical illness in people with SMI

Supplementing research on the direct consequences of mental distress arising from acute relapses in SMI, studies over a number of years have repeatedly shown that people with mental ill health die younger and a greater proportion have poor physical health, compared with the general population (Saha S 2007). Predictors of mortality and potential underlying pathways in relation to SMI will be more systematically reviewed in Chapter 2, but a broader overview of relevant issues and policy is provided here as an introduction.

Mental ill health does not occur in isolation, and it has been estimated that 46 per cent of people with a mental health condition also have a long-term physical health diagnosis (Naylor C et al 2012). In particular, those with mental ill health have high rates of respiratory, circulatory and infectious disease, obesity, abnormal lipids and diabetes (DE Hert M, Correll et al 2011; Pendlebury & Holt 2010; Bill et al. 2017; De Hert M et al 2009). A Finnish study estimated that up to 50 per cent of the increased mortality in those with mental ill health was due to underlying physical health conditions (Joukamaa M 2001). When it comes to disease-specific mortality, the death rates from respiratory disease and diseases of the digestive system are four times higher in mental health service users compared with the general population and 2.5 times higher for death from circulatory diseases (Dorning, H. Davies, A. Blunt 2015).

Drawing on information to date on these links between physical and mental health, a recent report from the Chief Medical Officer (Davies SC 2013) highlighted that approximately 60

per cent of the excess mortality in people with mental illness is avoidable. Considering physical health alongside mental health is imperative in order to address the disparities in life expectancy between those with and without mental ill health and one key area is to improve access to physical healthcare for those with mental ill health (Millard & Wessely 2014). NHS England suggests that there is an excess of 40,000 deaths among SMI patients, which could be reduced if those patients received the same healthcare interventions as the general population (NHS England 2015). For example, UK National Institute of Health and Care Excellence (NICE) guidance states that everyone with schizophrenia should have an annual physical health check (NICE 2014), yet a national audit in 2012 found that only 29 per cent of people with SMI received appropriate physical health checks (Cooper S et al, 2012). People with schizophrenia are also half as likely to undergo coronary revascularisation as patients without mental ill health (Davies SC 2013).

In the UK, at policy level, there has been greater focus at both national and local levels on achieving 'parity of esteem' for physical and mental health – to ensure that people are enabled to maintain both their physical and mental wellbeing (National Confidential Enquiry into Patient Outcome and Death (NCEPOD) 2017). Nationally, a Mental Health Taskforce was formed to develop a five-year national strategy for mental health, covering access, choice of treatments and prevention (NHS England 2015). These initiatives were accompanied by increased funding for mental health services, and the introduction of waiting time targets and quality standards for mental health services. In addition, a provider incentive payment was introduced to encourage secondary care mental health services to improve the physical healthcare of patients with mental health conditions. For these policy

measures to be effective, it is important to better understand how those with mental ill health are using services and whether they can access care for their physical health needs. A Health and Social Care Information Centre review found higher rates of access to hospital services for those with mental ill health compared with those without – but no distinction was made to compare hospital use for underlying physical or mental healthcare diagnoses (Health and Social Care Information Centre. 2013). Additionally, as the care services and pathways for physical and mental health are often not linked, data are captured in different ways in different systems, making it difficult to explore quality of care for physical health in those with mental ill health.

1.5 Wider health and healthcare inequalities

Among individuals with SMI, issues with help seeking behaviour and the resulting marginalisation of individuals is apparent. Patient/illness factors (DE Hert et al. 2011a; De Hert et al. 2011b) lead to difficulty in understanding health care advice combined with the motivation required to adopt new changes in lifestyle, poor compliance with treatment, cognitive deficits, reduced pain sensitivity (induced by antipsychotic medication)(Sendt et al. 2015). Poor communication and deficient social skills (e.g. as negative symptoms of schizophrenia) potentially account for the shortened life-span of patients with SMI(Kessing LV, Vradi E, McIntyre RS 2015; DE Hert M 2011b). An additional patient/illness factor is that psychiatric symptoms may render patients less inclined to discuss physical problems(Goodwin et al. 2016). Some doctors are uncomfortable dealing with psychiatric patients because the latter may be cognitively compromised which may impair or impede a doctor's clinical assessment. The stigma of mental illness, often the result of disparaging media coverage and negative stereotypes surrounding psychiatric patients, are other hurdles that prevent people from seeking treatment. Furthermore, people with mental disorders may be less likely to see a primary care physician and therefore to receive other interventions such as screening for cancer.

Unequal quality of healthcare provision could potentially explain much of the subsequent excess mortality(Goodwin et al. 2016). Psychiatrist-related factors are characterised by an over-emphasis on mental health to the exclusion of physical health, infrequent screening rates for metabolic abnormalities, omission of medical examination of patients because

physical complaints frequently are part of the psychiatric presentation, poor communication with the patient and the primary care teams, a lack of awareness and perhaps adherence to treatment guidelines, insufficient medical knowledge, and erroneous, sometimes misguided beliefs about patients' capability to change their lifestyle (DE Hert et al. 2011a; De Hert et al. 2011b). Factors common to the psychiatrist and other physicians include a tendency to dismiss or interpret physical symptoms as psychosomatic, lack of good quality care, unequipped teams, insufficient assessment, and difficulties providing consistent monitoring and continuity of care. Other physician-related factors relate to problems coordinating psychiatric and medical care.

In general, probably common to all suffers of mental illness, service-provision factors as outlined by NICE (NICE 2014)include a lack of clarity and consensus as to where the responsibility of physical health lies (DE Hert M, Correll CU, Bobes J, Cetkovich-Bakmas M, Cohen D, Asai I, Detraux J, Gautam S, Möller HJ, Ndetei DM, Newcomer JW, Uwakwe R 2011). The role of general practitioners in supervising healthcare for people with SMI who live in the community is variable. The fragmentation of medical and mental health care systems, lack of integration of services (poor or absent liaison links) and insufficient funds to resource the mental health service, limit the ability of most psychiatrists to focus beyond their own specialty. Service and system changes are prevalent in industrialised countries because reforms in mental health have led to reduced inpatient resources leading to shorter and infrequent hospital admissions with less time available to focus or investigate physical health problems. However, it is of concern when the highest number of excess deaths in schizophrenia is associated with cardiovascular disease, but still people with schizophrenia

have low rates of surgical interventions such as stenting and bypass grafting (Laursen TM, Munk-Olsen T, Agerbo E 2009; Lawrence D, Holman CDJ 2003). Similarly, individuals with psychosis are less likely to receive cerebrovascular arteriography or prescribed anticoagulants following stroke (Kisely S 2009), and of elderly patients hospitalised with heart failure, those with mental health problems receive poorer care during hospitalisation and have a greater risk of death and readmission (Rathore et al. 2008). In diabetes, patients with co-occurring mental health conditions are less likely to receive standard levels of care (Frayne 2006), and there are concerns about metabolic disorders going untreated in patients with schizophrenia, with non-white women especially vulnerable to under-treatment of dyslipidemia and diabetes (Nasrallah et al. 2006). Individuals with SMI are more likely to miss out on routine cancer screening (Carney & Jones 2006; Woodhead et al. 2016). Even a condition such as appendicitis has been found to present late and with more complications among individuals with schizophrenia and with worse outcomes from surgery (Cooke BK, Magas LT, Virgo KS, Feinberg B 2007). Meta-analyses of quality of medical care for people with co-morbid mental illness, have concluded that the majority of studies demonstrate significant inequalities in receipt or uptake (Mitchell AJ 2009). Chronic health problems are very common in patients with bipolar disorder, and tend to be under-recognised and sub-optimally treated (McIntyre RS et al. 2007). Most studies report on medical comorbidity in bipolar samples; relatively fewer studies report the reciprocal association. Individuals with bipolar disorder are differentially affected by several 'stress-sensitive' medical disorders notably circulatory disorders, obesity and diabetes mellitus (McIntyre RS et al. 2007) with iatrogenic factors and insufficient access to primary, preventive and integrated healthcare systems are all contributory.

1.6 Conclusion

Despite significant investments over the last 100 years, the mechanisms underlying the development and maintenance of psychosis remain unclear. While the emerging technologies of neuroscience and genetics have provided greater access to patients at the level of genes, protein, cells, and circuits, these new data have been connected only loosely to more clinically relevant psychopathology. People with the disorders underlying SMI experience substantial inequalities in many respects, with physical health outcomes becoming increasingly salient. Understanding relationships between the nature and characteristics of the mental disorders and associated health outcomes is important for developing interventions to address such inequalities, and the next chapter will review these relationships in more detail.

2 Literature review of mortality studies in people with severe mental illness

2.1 Chapter outline

This chapter aims to review all observational studies that have investigated predictors of mortality outcomes for people with schizophrenia, schizoaffective disorder and bipolar disorder and to consider methodological limitations of individual studies.

2.2 Objectives

The specific objectives of this review were-

- In the selected studies which investigate risk factors for mortality within SMI, to describe and compare all-cause mortality rates across SMI stratified by its three component conditions (schizophrenia, bipolar disorder, schizoaffective disorder)
- In the selected studies, to summarise reports of relative mortality rates against comparator groups including the general population
- To review reported predictors of mortality in people with SMI including particular consideration of symptom profile as a predictor of mortality

2.3 Methods

2.3.1 Search strategy

The search strategy incorporated the following databases: Embase and Ovid MEDLINE(R). The final search was carried out in May 2017 and is summarised in Table 2.1. Papers were initially selected by undertaking a search of articles on Medline. Relevant papers to SMI were identified by searching for key search terms on Table 2.1 which included the terms essential for this search including ‘schizophrenia’, ‘psychotic disorders’, ‘schizoaffective disorder’, ‘bipolar disorder’, ‘mood disorders’, ‘affective disorders’ and ‘depressive disorder’ where these overlapped with search terms for mortality. This search was supplemented with reviews of the bibliographies of relevant articles and from Google Scholar. SMI was defined and identified by a diagnosis of one of the component conditions- schizophrenia, bipolar disorder, schizoaffective disorder or of SMI as a broad category.

Table 2:1 - Search strategy for literature review

Electronic databases:

MEDLINE – January week 1, 2003 to May week 4, 2017, EMBASE – January, 2003 to May, 2017

Limits: Humans

English Language Search terms only

MeSH terms Medline:

Search ((((Schizophrenia[Title] OR Schizoaffective[Title] OR bipolar[Title] OR psychosis[Title] OR "antipsychotic agents"[Title] OR "antimanic agent"[Title]OR "Serious mental illness"[Title]))))) AND (mortality[Title] OR death[Title] OR "mortality rate"[Title] OR "SMR"[Title] OR "standa* mortality ratio"[Title])

Further restrictions applied-

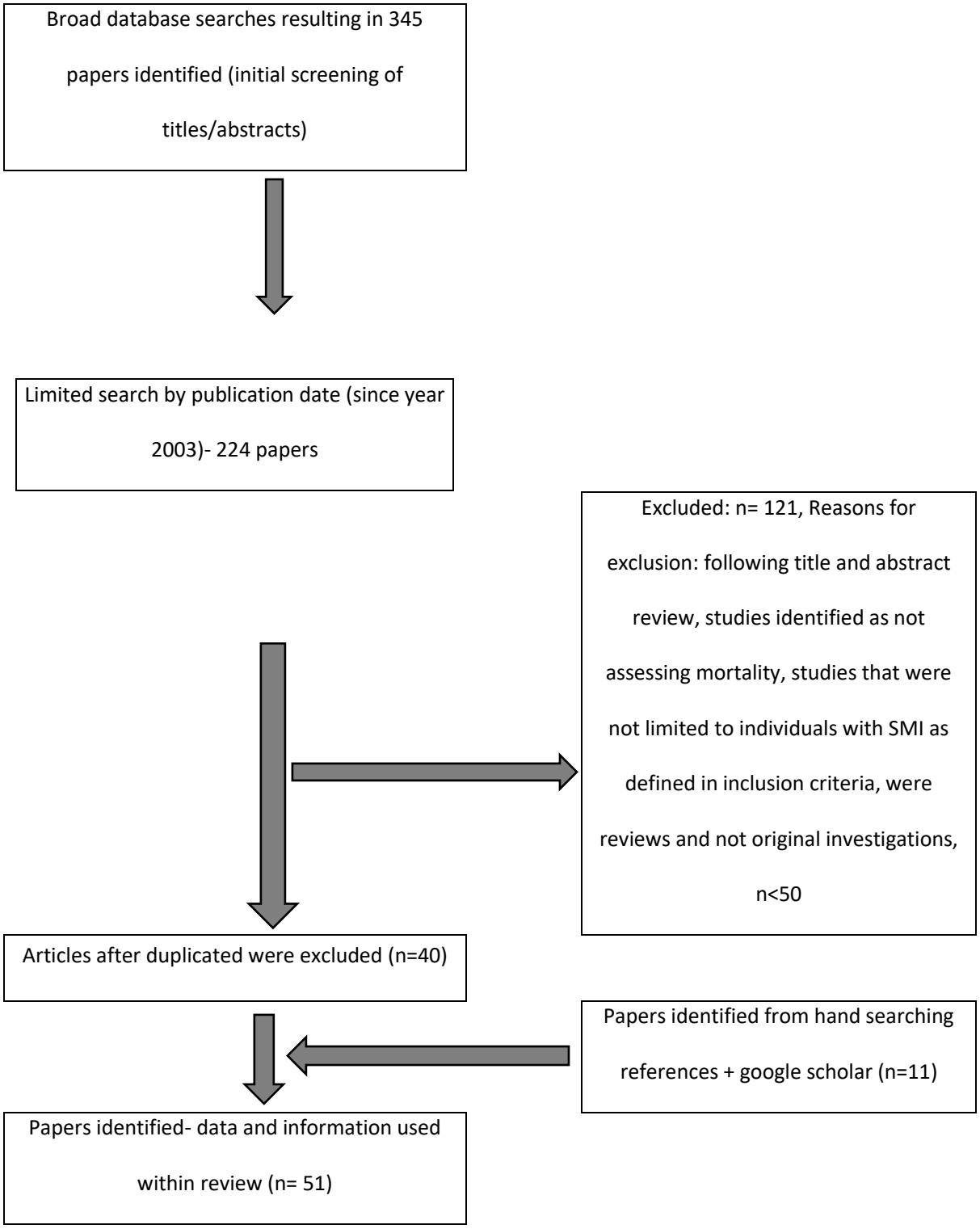
Search ((((Schizophrenia[Title] OR Schizoaffective[Title] OR bipolar[Title] OR psychosis[Title] OR "antipsychotic agents"[Title] OR "antimanic agent"[Title]OR "Serious mental illness"[Title]))))) AND (mortality[Title] OR death[Title] OR "mortality rate"[Title] OR "SMR"[Title] OR "standa* mortality ratio"[Title]) Filters: published in the last 10 years

Search ((((Schizophrenia[Title] OR Schizoaffective[Title] OR bipolar[Title] OR psychosis[Title] OR "antipsychotic agents"[Title] OR "antimanic agent"[Title]OR "Serious mental illness"[Title]))))) AND (mortality[Title] OR death[Title] OR "mortality rate"[Title] OR "SMR"[Title] OR "standa* mortality ratio"[Title]) Filters: published in the last 5 years

2.3.2 Study selection- Inclusion and exclusion criteria

Full text articles were retrieved to assess more thoroughly for inclusion for those which could not be excluded based on the criteria described below. Methodological strengths and weaknesses of different approaches were considered when interpreting and collating findings. For example, studies which mentioned collecting data from clinical records were included as cohort studies because it was assumed that clinical information would have been prospectively recorded. However, if in a study, data were collected and analysed as an ecological study and not at individual level, the study was excluded. Figure 2.1 shows the flowchart of search results and studies included in the final review.

Figure 2-1 - Flowchart of search results and studies included in final review



2.3.3 Inclusion criteria

This literature review included peer reviewed observational research. Studies were eligible for inclusion based on the following criteria:

i) published in English,

ii) published from 2003 onwards to provide review of papers published in the last 13 years,

iii) peer reviewed observational research,

iv) one of the following study designs – cohort/ data linkage study (prospective or retrospective), studies involving sample recruitment and collected data from individuals with SMI,

v) sample with SMI at least 50 in number.

2.3.4 Exclusion criteria

Papers identified by searches were excluded if they were conference abstracts or reports, letters, case reports or case series, ecological studies, reviews or qualitative studies and cohort studies of SMI with a sample size below 50. However, bibliographies of reviews were checked for additional references as described in Figure 2.1.

2.4 Results

2.4.1 Overview

For this review, 51 papers were identified as appropriate from the literature search. Table 2.2 summarises core characteristics of each included study including country, sample size, data source, study design & inclusion criteria. Most studies (n=41, 80%) included were from Europe or from the United States of America with the remainder from Australia, Canada and Asia. A variety of data sources were used in the different studies ranging from registry based data, hospital records, data linkages, and insurance databases in addition to data from bespoke cohorts. This resulted in a wide range of sample sizes from 72 (Morgan et al.,2003) to over one million (Olfsen et al., 2015). All studies were observational studies with cases either defined as SMI or schizophrenia or bipolar or schizoaffective disorders. In some studies, comorbid conditions were included, such as in the manuscript reported by Davis et al. (2012) describing patients with diabetes and schizophrenia. All studies used ICD-10 or DSM-IV classification systems.

Table 2:2- Description of studies from articles reviewed: predictors of mortality among patients with SMI

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Aagaard et al. 2016)	Denmark	1995 (n=1,100) 1998 (n=1,965) 2001 (n=2,327) 2004 (n=2,756) 2007 (n=2,725)	Danish Civil Registration System	Historical prospective record linkage study of patients with at least one visit to a Danish psychiatry emergency rooms in 1995–2007. Five consecutive 3-year cohorts of individuals aged 20 to <80 years included. Data from Danish Civil Registration System linked to Cause of Death Register & Central Psychiatric Research Register. Logistic predictor analyses for premature death performed.	SMI- during at least 2 years of primary diagnosis as F20, F22, F25 or F30–31 and at least four admissions or 50 inpatient days or secondary diagnosis as F1 or Z04.6
(Bowersox et al. 2012)	USA	3300	Dept. for Veteran affairs	Veterans with SMI lost to Veterans Affairs care for more than 1 year were contacted by providers who attempted treatment reengagement. Fisher's Exact Tests evaluated associations between mortality cause and	Veterans with SMI lost to Veterans Affairs care for more than 1 year

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
				reengagement status, and a Cox proportional hazard model evaluated the association between reengagement and survival.	
(Brown et al. 2010a)	UK	370	Hospital records	Cohort study- people with schizophrenia who had contact with local National Health Service psychiatric services between 1 January 1981 and 31 January 1982	Cases- 'firm diagnosis of schizophrenia' made by the responsible consultant; case-note evidence of first-rank symptoms of schizophrenia or persistent non-affective delusions or auditory hallucinations, in the absence of organic brain disease; and misuse of alcohol or other substance. Comparator- general population

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Brown et al. 2010b)		201	Social Security Administration's (SSA's) Death Master File (DMF) was used	Participants with SMI were recruited from six public and private outpatient mental health clinics in urban and suburban communities across the Baltimore metropolitan area to represent a broad range of patients receiving outpatient care. Participants with type 2 diabetes and without SMI were recruited from three primary care clinics near the psychiatric clinics.	Individuals with schizophrenia(n=201), major mood disorder(n=101) and persons without mental illness (n=99) in the prior year, all of whom had type 2 diabetes, were recruited from September 1999 until September 2002
(Brown S 2012)	UK	95	Hospital records	Cohort study- community cohort	Schizophrenia
(Callaghan et al. 2014).	USA	Schizophrenia= 174,277, bipolar= 78,739	Inpatient records were linked to death-certificate data	Cohort study- Age-, sex-, and race-adjusted standardized mortality ratios (SMRs) for the 19 diseases causally linked to tobacco use.	Individuals aged 35 or older at the mid-point of their follow-up period, with ICD-9 diagnoses of schizophrenia and related disorders
(Capasso et al. 2008)	USA	319	Medical records linkage system	Cohort study between 1950 and 1980 were followed until February 2005 for a median of 23.5 years	Residents meeting DSM-IV-TR criteria for schizophrenia or

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
					schizoaffective disorder. Comparator- general US population
(Castagnini et al. 2013)	Denmark	Bipolar= 3200, schizophrenia= 4576	Danish Psychiatric Central Register	Record linkage study to the official register of causes of death	All cases aged 15–64 years who were listed for the first time in the Danish Psychiatric Register between 1995 and 2008 with an ICD-10 diagnosis of bipolar disorder and schizophrenia. Deaths were identified by regular national tracing returns after 2006.
(Chang et al. 2010)	UK	38,066	SLaM BRC Case Register	Cohort study- included were schizophrenia (F20), schizoaffective disorders (F25), and bipolar affective disorder (F31). Substance use disorders (F10 to F19), depressive episode (F32) and recurrent depressive disorder (F33)	Standardized mortality ratios (SMRs) were calculated for the period 2007 to 2009 using SLaM records for this period and the expected number of deaths from age-specific mortality statistics for

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
					the England and Wales population in 2008(Chang et al. 2010).
(Crump et al. 2013).	Sweden	6618	Swedish Outpatient Registry, Swedish Hospital Registry, Swedish Pharmacy Registry	National cohort study- followed up for physical comorbidities	Bipolar disorder (ICD-10-F31),which was identified by any outpatient or in patient diagnosis during the preceding 2 years. All adults 20 years or older who had lived in Sweden for at least 2 years as of 2003 (N=6,587,036)
(Das-Munshi et al. 2017)	UK	18,201	SLaM BRC Case Register	longitudinal cohort study between 2007 and 2014. Linked mortality data from the UK Office for National Statistics for the general population in England and Wales. Cox proportional hazards regression to estimate hazard ratios and a modified Cox regression, considering competing	People with severe mental illness for five ethnic groups (white British, black Caribbean, black African, south Asian, and Irish)

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
				risks to derive sub-hazard ratios, for the association of ethnicity with all-cause and cause-specific mortality.	
(Davis et al. 2015).	Australia	1,291	Western Australia data linkage system	Longitudinal study	Type 2 Diabetics as cases. Comparators from electoral roll
(Dickerson et al. 2014)	USA	517	National Death Index	Prospective cohort study- assessed with clinical evaluation and blood sample from which antibodies to human herpes viruses and Toxoplasma gondii were measured	Schizophrenia patients vs. general population
(Dutta et al. 2007)	UK	239	Camberwell Case Register and hospital records	longitudinal study-indirectly standardized for age and gender. Univariate and multivariate analyses were used to test potential risk factors for suicide	All patients who presented for the first time with a DSM-IV diagnosis of bipolar I disorder in a defined area of southeast London over a 35-year period (1965–1999), Mortality rates were compared

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
					with those of the 1991 England and Wales population
(Ernst et al. 2013)	Denmark	14,974	Danish Psychiatric Research Register and the Danish Cause of Death Register	1980- 2010 Cohort study using linkage systems	Patients diagnosed with schizophrenia and representative sample of Danish population, excluding persons diagnosed with schizophrenia
(Fiedorowicz et al. 2009)	USA	435	Participants recruited for the National Institute of Mental Health Collaborative Depression Study	Prospective cohort study of Caucasian, English-speaking individuals knowledgeable about their biological parents, and were recruited between 1978 and 1981	comparison between participants with bipolar I and II disorder
(Fiedorowicz et al. 2014)	USA	1,716	Database of National Institute of Mental Health Genetics Initiative Bipolar Disorder	Cohort design- assessed the relationship between the duration of the most severe depressive and manic episodes and time to	Final best estimate diagnosis of bipolar I disorder from centres that due to the wording of the original informed consents provided

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
			Consortium and National Death Index	vascular mortality (cardiovascular or cerebrovascular)	personal identifiers to facilitate searching of the National Death Index
(Fors et al. 2016)	Sweden	255	National population register	Retrospective observational study 1981-1991	All persons with schizophrenia diagnosed per DSM-III-R criteria identified for the period
(Hayes et al. 2012a).	UK	6880 SMI (4064, 745, 1770- Schizophrenia, Schizoaffective, Bipolar)	Case register linked to national mortality tracing	This dynamic cohort consisted of individuals who were diagnosed with SMI during a specific observation period (from the 1st of January 2007 to the 31st of December 2010, inclusive) and who had been assessed by a clinician using the Health of the Nations Outcome Scale (HoNOS), at least once during this observation period- symptoms.	Patients were classified as having an SMI if, while in contact with Mental Health hospital services and they had been diagnosed with schizophrenia (ICD-10 code: F20), schizoaffective disorder (F25) or bipolar affective disorder (F31) during the observation period

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Hayes et al. 2012b).	UK	6,880 SMI (4064, 745, 1770- Schizophrenia, Schizoaffective, Bipolar)	Case register linked to national mortality tracing	Cohort study- individuals diagnosed with SMI from the 1st of January 2007 to the 31st of December 2010 and assessed by a clinician using the Health of the Nations Outcome Scale (HoNOS), at least once during this observation period- for functional status.	Patients were classified as having an SMI if, while in contact with Mental Health hospital services and they had been diagnosed with schizophrenia (ICD-10 code: F20), schizoaffective disorder (F25) or bipolar affective disorder (F31) during the observation period
(Hayes et al. 2015).	UK	14,754	Case register linked to national mortality tracing	Cohort study modelling the effect of clozapine on mortality over a 5-year period (2007-2011) using Cox regression	SMI diagnosis (WHO ICD-10 codes: F20, F25, F31) during 2007 to 2011. Analyses compared individuals who were newly prescribed clozapine against those with no evidence of this agent being prescribed (excluded if already had clozapine in the past)

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Hjorthøj et al. 2015).	Denmark	Schizophrenia= 41,470, Bipolar= 11,739	Linked nationwide registers	Prospective, register-based cohort study, treatment for substance use disorders (categorised into treatment for alcohol, cannabis, or hard drug misuse), date of death, primary cause of death, and education level.	Cases included unipolar depression in addition to SMIs. Comparators from Danish civil registration system
(Hoang et al. 2011)	UK	373,099 discharges	English hospital episode statistics and death registration data	Record linkage study	People discharged from inpatient care with a diagnosis of schizophrenia or bipolar disorder, followed for a year after discharge.
(Hoang et al. 2013)	UK	51,624	English national Hospital Episode Statistics and death registrations from Office for National Statistics	Prospective record linkage study	people aged <75 years, discharged from inpatient care with schizophrenia or bipolar disorder in 2006– 2007, and followed up for 365 days.

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Høye et al. 2011)	Norway	1,111	Case register of hospital linked to the Causes of Death Register of Norway	Cohort study- Schizophrenia was defined per the International Classification of Diseases (ICD-9:295, ICD-10: F20, F21and F25)	Observed from the date of first admission after January 1st 1980 until death, moving abroad or the end of 2006, for a total of 16129 person years. In-hospital patients at January 1st 1980 were included, with start of follow-up the same day. All death certificates are coded by a physician per the current ICD- system.
(Kim et al. 2017)	South Korea	Schizophrenia= 9,387	National Health Insurance	Cohort study between 2002 to 2013- Poisson regression modelling conducted to quantify the effect of baseline characteristics on all-cause and suicide mortality risks.	Mood disorder, or mental and behavioural disorder due to psychoactive substance use and to compare this with that of the general population included in addition to schizophrenia

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Kisely et al. 2013)	Western Australia	2958	Mental Health Information System of psychiatric inpatient, outpatient and community contacts, Mental Health Review Board database of compulsory psychiatric treatment, Hospital Morbidity Data System; and mortality data	Record linkage observational study- Cohort included patients from all community-based and inpatient psychiatric services	Compulsory community treatment, such as community treatment orders
(Kredentser et al. 2014)	Canada	9,038	De-identified administrative databases	Population-based analysis was performed to compare age- and sex-adjusted 10-year (1999-2008) mortality rates, overall and by specific cause, of decedents aged 10 years or older who had 1 diagnosis of schizophrenia (ICD-9-CM code 295, ICD-10-CA codes F20, F21, F23.2, F25) over a 12-year period	Includes F21, F23. Comparator is general population

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Lahti et al. 2012)	Finland	12,939	Helsinki Birth cohort study	Cohort study- with 30 years follow up by using national databases on cardio- and cerebrovascular hospitalizations and mortality and on reimbursement entitlements and use of drugs for treatment of hypertension, dyslipidaemia, coronary heart disease and diabetes	Cases limited to individuals with schizophrenia
(Laursen et al. 2011)	Denmark	16,079	Danish Civil Registration System	Cohort study. Patients with schizophrenia (ICD8: 295 (exclusive 295.79), ICD10: F20) and bipolar disorder (ICD8: 296.39, 296.19, ICD10: F30, F31) and with any of the remaining psychiatric diagnoses registered at psychiatric contact were identified	investigates and evaluates the impact of 19 severe chronic diseases on excess mortality due to diseases and medical conditions (natural death) in individuals with psychotic disorders
(Laursen et al. 2014)	Denmark	1,061,532	Danish population register, Danish Psychiatric Central Register	Prospective cohort study	Individuals with schizophrenia or bipolar disorder with subjects with no prior psychiatric hospitalisation

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Medici 2015)	Denmark	15,334	Danish Psychiatric Central Research Register, Danish Register of Causes of Death	Cohort study- bipolar disorder (International Classification of Diseases-10: F31)	First-ever diagnoses of bipolar disorder (International Classification of Diseases-10: F31) between 1995 and 2012. Age- and gender standardized incidence rates, standardized mortality ratio (SMR) and Kaplan-Meier survival estimates
(Morden et al. 2012)	USA	65,362	Veteran Affairs health system	Matched cohort study- annual prevalence of diagnosed cardiovascular disease, diabetes, dyslipidemia, hypertension, obesity, and all-cause and cause-specific mortality was compared for fiscal years 2000-2007. Mean years of potential life lost (YPLLs) were calculated annually	65,362 patients in the Veteran Affairs (VA) health system with schizophrenia to 65,362 VA patients without serious mental illness (non-SMI) matched on age, service access year and location

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Morgan et al. 2003)	Ireland	72	Hospital records and interviews	Cohort study- All persons with schizophrenia, including both inpatients and those living in the community,	Individuals with schizophrenia followed up 7.5 years after case ascertainment
(Olfson et al. 2015)	USA	1,138,853	Medicaid program, Death record information was obtained from the National Death Index	National retrospective longitudinal cohort of patients with schizophrenia 20 to 64 years old during 2001 to 2007.	Adults with schizophrenia
(Ösby et al. 2016)	Sweden	614 035 person-years	National health registers	Population-based register study including all Swedish patients in adult psychiatry admitted to hospital with a main diagnosis of schizophrenia, bipolar or unipolar mood disorder in 1987–2010	Mortality rates adjusted for age, sex and period were calculated using direct standardization methods with the 2010 Swedish population as standard. Data on all residents aged 15 years or older were used as the comparison group(Ösby et al. 2016)(Ösby et al.

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Pan et al. 2016a).	Taiwan	7987	Health Insurance system	Cohort study	2016)(Ösby et al. 2016)(Ösby et al. 2016)(Ösby et al. 2016)(Ösby et al. 2016)(Ösby et al. 2016)(Ösby et al. 2016)(Ösby et al. 2016)(Ösby et al. 2016) Cases- diagnosed with bipolar in 2008 aged 18 years or above with data were available for a minimum of 12 months before and 36 months after the index date and did not have a diagnosis of bipolar in the year preceding the index date
(Pan et al. 2016b).	Taiwan	15,254	National Health Insurance Research Database	Cohort study of newly diagnosed bipolar disorder	Survival analyses were performed with this national cohort to examine the associations between the first-year treatment pattern

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
					(hospitalisation and number of outpatient clinic visits) and mortality over a follow-up period of 3 years (2008-2011)
(Pan et al. 2017)	Taiwan	272,617	National Health Insurance Research Database	Cohort study to estimate age- and gender-standardized mortality ratios (SMRs) were calculated for each of the 3-year observation periods. Weight gain and metabolic syndrome in relation to use of second-generation antipsychotics SGAs assessed	Two nationwide samples of individuals with schizophrenia and bipolar disorder
(Ramsey et al. 2013)	USA	14,870	Household survey, institutional survey, structured psychiatric diagnostic interview	Retrospective case control study of bipolar disorder- 1) manic episode (n=46); 2) hypomanic episode (n =195); 3) subthreshold manic symptoms (n =1,041); and 4) no manic spectrum episodes (n =13,588).	Participants with manic spectrum episodes were compared to those without such episodes for mortality after 26 years
(Reininghaus et al. 2015)	UK	557	Mental health services records	First-episode psychosis initially identified in 2 areas (southeast London and Nottinghamshire,	Cohort of individuals with SMI

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
				United Kingdom) were traced over a 10-year period in the ÆSOP-10 study	
(Schoepf D, Uppal H, Potluri R 2014)	UK	1418	Local health authority computerized hospital activity analysis register	Retrospective case–control study with up to 12-year follow-up in general hospital admissions from 2000 to 2012	Cases were individuals with schizophrenia. 14,180 age- and gender-matched hospital controls
(Talaslahti et al. 2012)	Finland	9461	Nationwide registers	Cohort study- Patients retired because of schizophrenia or schizoaffective disorder also included. Those living abroad or in the Åland Islands were excluded. Mortality and causes of death of this sample were followed up using the registers for the 10-year period (1999–2008).	Individuals with schizophrenia over 65 years by the first of January 1999, with schizophrenia or schizoaffective disorder (ICD-8, ICD-9: 295, ICD-10: F20, F25) as the main register diagnosis during the period 1969–1998
(Talaslahti et al. 2015)	Finland	6,142	Nationwide registers	Cohort study with register based follow-up for mortality covered 10 years between 1999 and 2008	Cases- individuals with schizophrenia aged 65years or above, mortality and causes of death diagnosed at age of 60+ (very-late-onset schizophrenia-like

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
					psychosis, VLOSLP) Comparator- sex- and age-matched general Finnish population. Comparator- earlier onset (below 60years) schizophrenia patients
(Tenback. et al. 2012)	USA	7,415	Health insurer database	Observational study examining schizophrenia diagnosis and mortality. Includes medication, antipsychotic use, type of antipsychotic, other medication	Patients with schizophrenia using an antipsychotic (n = 7415) and a group of randomly chosen control subjects (n = 97,726)
(Tiihonen et al. 2016)	Sweden	21,492	National Patient Register &MiDAS register (maintained by the Social Insurance Agency of Sweden), Prescribed Drug Register	Prospective population-based cohort study of patients with schizophrenia aged 16–65 years of age	Cumulative low, moderate, and high exposure to antipsychotics, antidepressants, and benzodiazepines from 2006 through 2010. all antipsychotics dispensed from 2006 through 2010 except lithium. cumulative exposure of antipsychotics was

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
					estimated by using the defined daily dose
(Toffol et al. 2015).	Finland	826	Hospital Discharge Register	All the individuals who were hospitalized in Finland because of a suicide attempt (International Classification of Diseases, 10th Revision ICD-10 codes X600 to X840, Z728, or Z915) between January 1, 1996 and December 31, 2003, and in prospective screening had been hospitalized due to bipolar disorder (ICD-10 codes F30.0 to F30.9, F31.0 to F31.9) before the index attempt, were included in the study	Relative risk of suicide attempts leading to hospitalisation, completed suicide, and overall mortality during lithium vs. no-lithium, antipsychotic vs. no-antipsychotic, valproic acid vs. no-valproic acid, antidepressant vs. no-antidepressant and benzodiazepine vs. no-benzodiazepine treatment was measured.

Study	Country	Sample size	Data source	Study design & inclusion criteria	Case+/- comparator definitions
(Westman et al. 2013)	Sweden	17 101	Swedish Total Population Register Swedish Cause-of-Death Register and the National Patient Register	Cohort study with a 20-year follow-up of persons were diagnosed with bipolar disorder between 1987 and 2006.	Bipolar diagnoses recorded between 1987 and 1996 were identified using ICD-9 (296A, C, E and 298B) from 1997 and ICD-10 (F30-F31).
(Wu et al. 2015)	Taiwan	11,247	National Health Insurance Research Database	Population-based matched cohort study- Cox proportional hazard models to determine the effect of schizophrenia on macrovascular and microvascular complications, and all-cause mortality	11 247 participants with diabetes and schizophrenia and 11 247 participants with diabetes but not schizophrenia were enrolled

2.4.2 Types of studies included-Clinical records based studies

Overall, 22 studies utilising clinical records as their primary data source investigated all-cause mortality or factor mortality such as age, gender or ethnicity. The sample sizes (individuals with schizophrenia) varied from less than 100 individuals in one study to over a million individuals with schizophrenia, bipolar disorder or schizoaffective disorder. The strengths and the weaknesses of these studies are summarised below.

Callaghan et al. 2014 investigated outcomes for age, sex, and race-adjusted standardised mortality ratios among 253,016 individuals with schizophrenia and bipolar disorder. This study was large but its limitations included some methodological considerations in case-ascertainment. For example, tobacco-related cancer deaths might as a result be partly misattributed to other causes commonly occurring such as cardiovascular problems or more easily apparent conditions such as bronchitis or emphysema present near the time of death. In addition, competing risks of death from other causes, especially those occurring earlier in the lifespan than tobacco- related cancers, might also play a role. The use of inpatient diagnoses of schizophrenia, bipolar disorder, and depression means that the study included only individuals who both required and received inpatient care. This will have resulted in a sample with relatively severe illness, and may have excluded individuals with limited access to care or only received outpatient care. The available administrative medical records did not contain any information about the presence, frequency, intensity, or duration of tobacco use. Mortality estimates will be biased downwards to some extent due to out-

migration, as individuals leaving the state would be lost to follow-up but would continue to accrue time at risk. As the sample was restricted to individuals hospitalised with diagnosed psychiatric conditions, findings might be less generalisable to other psychiatric samples, such as those drawn from the general population or outpatient care. The study was not able to assess the validity of the psychiatric diagnoses by reviewing patient medical charts and, as a result, it is possible that such misdiagnoses biases may have affected the mortality estimates.

Chang et al. 2010 used a case register to investigate outcomes for all-cause mortality among 38,066 individuals with schizophrenia, schizoaffective disorder, bipolar affective disorder as well as depressive episode and recurrent depressive disorder. Case registers derived from secondary healthcare offer advantages for investigating “high penetrance” disorders where the chances of secondary care contact are high. Prevalence bias, however, is an issue which needs consideration, in that the cases known to a service within a given time are likely to be dominated by those with long and relapsing clinical courses - they therefore cannot be taken to generalise to incident cases. A further methodological challenge was that service data were principally (but not entirely) derived from a single catchment area whereas expected deaths were derived from national data. However, the local data were reflective of an urban city population, so retain at least some generalisability to similar populations. Diagnostic categories overlapped since a proportion of individuals suffer from more than one mental disorder, and confounders other than age and gender might still exist.

Crump et al. 2013 investigated all-cause mortality in 6,618 individuals with bipolar disorder. The study's limitations include possible confounding by disease severity or unmeasured factors, the fact that few individuals received monotherapy, and that precision was limited for specific medications. Unlike the study described by Callaghan et al. (2014), the availability of outpatient diagnoses is a strength because it allows the inclusion of milder, non-hospitalised co- morbidities. This reduces the selection effects that arise from the sole use of inpatient data and enables more generalisable risk estimates. In addition, individual data about smoking, exercise, or other direct lifestyle measurements were unavailable. Other substance use disorders were examined using nationwide outpatient and inpatient diagnoses, but incomplete ascertainment may have resulted in under-estimation of effects.

Das-Munshi et al. 2017 investigated outcomes for all-cause and cause-specific mortality by ethnicity among 18,201 individuals with schizophrenia, schizoaffective disorders and bipolar disorders. The study's strengths include use of a large cohort from an ethnically diverse location to assess differences in mortality outcomes for each ethnic minority group. By prospectively assessing mortality and tracing emigrations out of the cohort, the possibility of bias through emigration was addressed. A surveillance period of up to 8 years was carried out enabling more comprehensive ascertainment of patients in contact with the mental health trust with severe mental illness. The study was not able to capture people with psychosis who sought private health care but numbers were expected to be small in a UK context. The findings are generalisable by location to other urban inner cities where ethnic minority communities tend to reside. The study's limitations include the fact that diagnoses of severe mental illness were not based on research diagnostic criteria. Biased diagnostic

practices might have led to ethnic minority groups being more likely to receive a psychosis diagnosis that would not meet research diagnostic criteria, and thus lower recorded mortality risk if such misdiagnosis in ethnic minority groups meant that these individuals had less severe mental illness. In addition, the broad 'south Asian' category applied in this study might have masked important differences for Indian, Pakistani, and Bangladeshi individuals. The study adjusted for neighbourhood-level deprivation, but did not have data on individual-level socio economic position. The role of mediators of premature mortality, particularly tobacco use, type 2 diabetes, obesity, and hypertension were not addressed.

Hayes et al. 2012a investigated outcomes for all-cause mortality among 6,880 individuals with schizophrenia, schizoaffective and bipolar disorder. The sample was large and inclusive of all individuals with SMI who had made contact with mental health services within a defined area over a 4-year period. The mental health service is a large provider of secondary mental healthcare and a near-monopoly provider for its geographic catchment, hence the data should be representative of individuals with SMI living in urban and suburban areas. The longitudinal analysis utilised complete electronic clinical records of more than six thousand cases, providing the statistical power to simultaneously control for a range of potential confounders. In the UK, it is a legal requirement for primary and secondary healthcare providers to keep death records up to date. Mortality tracing in the source records system is updated monthly and is based on national certification so that only deaths occurring outside the UK are likely to have been missed. However, the issue of residual confounding remains, in particular, for factors such as medication use that were not included in this analysis. Also, a reliable assessment of duration of illness was not available

for this cohort as well as whether it was chronic or episodic illnesses. It was not possible to manually code free text fields in patient records due to the large sample size and numerous records per individual. Consequently, symptom assessment was restricted to Health of the Nation Outcome Scales (HoNOS) items. The HoNOS scales were developed to measure the health and social functioning of people with severe mental illness(Wing JK, et al 1998). In the HoNOS instrument, hallucinations and delusions are combined as one item making it impossible to analyse these separately, and there are no HoNOS scales covering symptom domains such as negative or disorganisation. Due to small numbers in some response categories, it was necessary to group together people who had mild through to very severe problems. It is possible that this grouping might have obscured associations. Furthermore, symptom assessment was based on the first HoNOS measurement in the observation period, and fluctuations in symptoms over time would not be well captured by a single HoNOS assessment. It is possible that mortality risk is mediated by other symptoms not included in this study, and the levels of disadvantage or referral bias may still influence the characteristics of the cohort who present to secondary care.

Hayes et al. 2012b reported mortality outcomes by functional status for the same cohort described above among 6880 individuals with schizophrenia, schizoaffective and bipolar disorder. Data on adverse lifestyle choices (other than drinking problems), such as smoking, poor diet, and physical inactivity variables were not available, although these are more likely to represent mediating rather than confounding factors for associations between a mental disorder and mortality. ADL impairment, one of the key exposures of interest, was assessed using a single HoNOS item, consequently it was not possible to explore the relative

contributions of basic versus instrumental ADLs. In addition, cognitive function was not assessed in this analysis. Three quarters of patients with SMI in the cohort had received at least one HoNOS assessment during the observation period but there were differences between those with and without this assessment. Most notably, those missing the HoNOS assessment were older and more likely to have died during the observation period, differences which may have impacted on results. It should be noted that HoNOS is not a research assessment, and inter-rater variation may have reduced the apparent effect of the exposures examined. All secondary mental healthcare within the UK National Health Service (NHS) is provided at no cost to consumers, so the only missing mental health service contacts would be from individuals seeking exclusively private healthcare. However, the characteristics of the cohort who are known to secondary care may still be influenced by levels of disadvantage or referral bias. Consequently, the generalisability of both findings is principally to secondary care rather than primary care populations but given the high penetrance nature of SMI described by Chang et al., 2010, secondary contact is expected from most individuals.

Hayes et al. 2015 investigated outcomes for premature mortality from natural causes among 14,754 individuals with schizophrenia, schizoaffective and bipolar disorder, specifically focusing on clozapine use as an exposure. This investigation has a number of strengths. The sample included all individuals with SMI in contact with mental health services within a defined area over a 5-year period. The study drew on complete electronic clinical records for close to fifteen thousand cases, providing the statistical power to control for a range of potential confounders. Its limitations include the fact that it only examined

cases newly prescribed clozapine during the observation period and did not investigate mortality associations beyond 5 years. Despite using propensity scores to perform sensitivity analysis and adjustment sensitivity, confounding by indication is an important consideration which cannot be ruled out entirely in any observational study. A strength is that the study adjusted for indicators of severity of illness including diagnosis, symptoms, physical illness, and functional status but not an assessment of duration of psychiatric illness was not available.

Hoang et al. 2011 investigated trends in age-standardised mortality ratios over time, comparing the mortality in individuals with schizophrenia or bipolar disorder with mortality in the general population, using data on 373,099 discharges among individuals with recorded diagnoses of schizophrenia or bipolar disorder. The strengths of this study include its large size and the fact that it includes all records of discharges with a diagnosis of schizophrenia or bipolar disorder in the study period and is nationally representative for England. By linkage to national mortality data there is comprehensive information on deaths, especially in the period after discharge from hospital when the risk is thought to be greatest. Potential limitations include the use of information from hospital records and death certificates, which were not collected for research purposes. There are also caveats about the validity of a dichotomy between “natural” and “unnatural” categories of death, although the definitions of these categories from the ICD codes were standard and applied consistently over time. The case groups consisted of people who had been recently admitted to hospital with a principal diagnosis of a mental disorder during the period in question and, similar to the sample used by Callaghan et al. (2014) are likely to represent a

relatively severely affected group at a period of higher risk. The severity of the disorders in the case groups might have changed over the monitoring period because of changes in admission policy. Another limitation as described before with other observational studies is that residual confounding with severity of illness cannot be excluded entirely.

Hoang et al. 2013 investigated outcomes for all-cause SMRs- natural/ unnatural among 51,624 individuals with schizophrenia or bipolar disorder. Its strengths include a large sample size of this study, and includes all records of discharges with a diagnosis of schizophrenia or bipolar disorder in the study period nationally, study characteristics aid the generalisability. A study limitation is the reliance on the accuracy of routinely collected data, which are not primarily collected for research purposes. The study only focused on investigating avoidable mortality risk in people who had been admitted with a diagnosis of schizophrenia or bipolar disorder which may have over-estimated the risk from these conditions as it did not include people who had not been recently admitted. Age and sex differences on avoidable causes of death examined but other determinants such as ethnicity and duration of mental illness, were not examined.

Kisely et al. 2013 investigated outcomes for 2-year all-cause mortality among 2,958 individuals with schizophrenia and affective disorder. This included psychiatric and physical morbidity measured through records of individuals' use of health services. Potential confounders include lack of insight or treatment adherence, social disability, types of medication and characteristics of the treating teams or services.

Kredentser et al. 2014 investigated all-cause mortality among 9,038 individuals with schizophrenia. The strengths of the study include its population-based analysis, accounting for the underlying general mortality rates and prevalence of diagnoses. The "cause of death" coding through the Manitoba government's Department of Vital Statistics was limited to a single primary cause of death. If the primary cause of death is not specific or misleading such as "death by mental illness", having additional coded causes of death would be useful. The death certificate data source also contains data from local residents only, which may limit generalisability of findings. In this study there was also limited ability to assess the various determinants of the causes of mortality often implicated in schizophrenia, such as smoking behaviour, weight, and comorbid illness symptoms such as blood-sugar levels.

Olfson et al. 2015 investigated outcomes for all-cause and cause-specific mortality among 1,138,853 individuals with schizophrenia. The main strength of this study is the large sample size; however, the study had no methods of validating the accuracy of schizophrenia diagnoses or estimating the proportion of cases ascertained in the source Medicaid claims data, and there is a possibility that different results would have been obtained if privately insured and uninsured individuals with schizophrenia were included in the analysis. The authors were also not able to investigate smoking status, body mass index, and substance abuse. In addition, in common with all studies investigating externally caused mortality, the impact of stigma is difficult to quantify and it is possible that it may have contributed to under-reporting of suicide and inflated accidental deaths. A further limitation could come

from Medicaid encounter data providing incomplete coverage of individuals enrolled in some managed care plans.

Schoepf D et al 2014 investigated general hospital mortality outcomes among 1,418 individuals with schizophrenia, and a hospital-based control group of 14,180 control subjects, matching each schizophrenia case with ten randomly selected hospital controls of the same age and gender which increases power of study. However, there were limitations to note. Firstly, due to matching method, the differences of ethnicity and marital status not measured. Secondly, discharge records and general hospital mortality data did not permit to differentiate between primary diagnoses and secondary or tertiary diagnoses. The possible effects of diagnostic awareness may have influenced findings. The awareness bias may have led to systematic under- reporting of physical diseases in the schizophrenia population and a systematic underreporting of mental diseases in the control population. For example, cardiovascular diseases such as hypertension, angina, and hyperlipidaemia may have been underreported in the schizophrenia population compared to the control population.

Davis et al. 2015 investigated outcomes for mortality among 1,291 individuals with schizophrenia and diabetes. There is possibility that there is healthy participant effect as questionnaire data were collected; however, the study was potentially informative through the comprehensive assessments carried out in a large community-based, representative sample and the well validated tools used for data from matched non-diabetic residents.

All of the remaining studies described here under clinical records study type have sample sizes less than 1,000, some even less than 100. They therefore contribute less substantially to the evidence base on the basis of statistical power, although often have advantages in sample selection and data availability. Brown et al. 2010a investigated 25-year mortality in people with schizophrenia with a particular focus on changes over time among 370 people with schizophrenia. The study was limited by missing people with unrecognised schizophrenia and those who avoided service contact during the index year, but was representative of Southampton residents with schizophrenia living outside hospital and known to local mental health services in 1981–82. The use of a cohort recruited 25 years ago means their experiences may differ from that of people diagnosed more recently. However, ascertainment was relatively rigorous, as status was independently checked and loss to follow-up was low. Brown et al., 2012 investigated outcomes for natural mortality among 95 people with schizophrenia. This study's strengths include its prospective design, face to face assessment and completeness of follow-up. The original diagnoses of schizophrenia were made by consultant psychiatrists based on the standard clinical criteria-individuals were representative of middle aged people with schizophrenia, living in the community in the UK making the study findings generalisable. There is, however, selection bias as the results are derived from people who were survivors of an initial cohort and who consented to interview.

Morgan et al. 2003 investigated outcomes for all-cause mortality among 72 individuals with schizophrenia. Its strengths include an epidemiologically complete cohort, and as such encompasses the totality of schizophrenia, from young, first episode cases, through a

majority of cases receiving care in the community, to a minority of elderly cases who continue to receive long-term inpatient care. There is still potential confounding factors such as residential mobility and socioeconomic, ethnic and geographical diversity but all cases were ascertained, interviewed in person and diagnosed using contemporary operational criteria by the same investigatory team, and the study utilised limited information was available on factors such as diet, smoking habits and exercise. Capasso et al. 2008 investigated outcomes for all-cause mortality among 319 patients with schizophrenia or schizoaffective disorder. The limitations include diagnosing retrospectively from the unified medical record leading to inherent limitations compared with prospective identification. Changes in clinical practice over time in history taking and documentation meant certain aspects of the medical and psychiatric interview were not recorded consistently, particularly smoking history. Dickerson et al. 2014 investigated all-cause mortality among 517 individuals with schizophrenia with strengths including detailed clinical assessment at baseline and the inclusion of blood-based antibody markers; however, information was missing on factors such as obesity and physical inactivity, and on other potentially relevant blood- based markers such as C-reactive protein, lipid levels, and blood glucose.

Dutta et al. 2007 investigated suicide and all-cause mortality rates among 239 individuals with bipolar I disorder. It is unlikely that the rating of risk factors was biased because this was done to create a case register and not to test specific hypotheses. However, the risk factors were only assessed using the specific criteria for the period until the first year after presentation, and data on suicide attempts during the follow-up period were not available.

If the recording of suicide on death certificates was more likely if the deceased was known to have had recent contact with psychiatric services, then this would have been a source of bias towards higher mortality ratios for suicide. Reininghaus et al. 2015 investigated outcomes for all-cause, natural-cause, unnatural cause mortality among 557 individuals with first-episode psychosis. However, the limited follow-up of this cohort may have also led to type 1 error and there is possibility of unmeasured confounding by other factors such as smoking, obesity and other unhealthy lifestyle factors, medication, and access to, or quality of, healthcare. This includes alcohol use at baseline because if measured this would have enabled, as for baseline illicit drug use, to investigate its role as a risk factor. Toffol et al. 2015 investigated all-cause and suicide mortality among 826 individuals with bipolar disorder. The study included only individuals with high-risk bipolar disorder who were hospitalised due to a suicide attempt but was nationwide from the whole Finnish population, and had a relatively long follow-up period. The study was only able to identify hospital-treated suicide attempts and the diagnosis of bipolar disorder was not based on standardised diagnostic interview schedules, but on a clinical diagnosis made by a physician.

2.4.3 Register based studies

Register-based research considered in this review will be research that has been based on data contained in records kept by government agencies or other organisations, and where data can be traced to individuals which could be for the entire population or individuals with a particular characteristic. Overall, 24 register based studies investigated all-cause mortality. Study size varied from less than 100 individuals in one study to over a million of individuals. The study strengths and weaknesses are summarised below.

Aagaard et al. 2016 investigated causes of death and mortality rates among individuals with severe mental illness. The limitations of this study include that its sample was drawn from users of psychiatric emergency rooms in a specific Danish context and, therefore, the results cannot necessarily be generalised to users of other mental health-care settings. In addition, the number of deaths from specific causes was too small in some cohorts for reliable assessment of trends over time and in general. Furthermore, local categorisation practice differences may make the register-based clinical data quality variable.

Brown et al. 2010b investigated all-cause mortality in 201 individuals with type 2 diabetes and SMI. Strengths include the relatively large cohort with this combination of conditions, who were extensively characterised at baseline and compared to a local group of individuals with type 2 diabetes who did not have SMI. Study limitations include the fact that causes of mortality were unknown, as was the extent to which mortality was attributable to diabetes.

Overall, the sample size and number of deaths limited the number of predictors of mortality that could be investigated in the multivariate logistic regression model.

Castagnini et al. 2013 investigated mortality and cause of death among 7,776 individuals with bipolar disorder and schizophrenia. The limitations of this study are those inherent to record linkage surveys which use information collected routinely from psychiatric and death registers.

Ernst et al. 2013 investigated trends in age of death among individuals with schizophrenia. The authors noted that findings may not be generalisable to countries with a different health system or social welfare; however, less ready access to healthcare would probably contribute to even greater differences between the groups of interest over time. Data on specific causes of death should be interpreted with caution, as the reliability of these estimates might vary over time and in quality. The main strengths of the study were the long follow-up and nationwide sample, in which data were gathered systematically and individuals not lost to follow-up. Primary and secondary care in Denmark is free at the point of access, removing a potential confounding factor compared with studies based in countries where free healthcare is less available.

Fiedorowicz et al. 2009 investigated cardiovascular mortality risk in 288 individuals with bipolar I disorder and 147 individuals with bipolar II disorder. Some participants were lost to follow-up prior to the assessment of the primary outcome, potentially leading to bias, and

this loss to follow-up limited the ability to assess exposure to medications and burden of clinically significant affective symptomatology. It may also overestimate the burden of affective symptoms and perhaps medication exposure for those lost to follow-up early, resulting in a differential misclassification of exposure. Limited observed exposure to several medication classes impeded the pursuit of a more refined medication exposure assessment, for example of more recently marketed or less commonly prescribed medications, limited instead to first- generation antipsychotics, lithium and specific anti-depressants. The treatment exposure was not controlled and may therefore be confounded by a variety of factors, including severity of illness and general health-seeking behaviours. The baseline cardiovascular risk assessment was limited by available data, for example not including dyslipidaemia, body mass index, and smoking history. The primary outcome variable, cardiovascular mortality was observed in 33 participants, limiting statistical power.

Fiedorowicz et al. 2014 investigated time to vascular mortality using Cox proportional hazards models in 1,716 participants with bipolar I disorder. Limitations include potential misclassification of exposure, given prior findings suggestive that manic symptom burden may be over-estimated retrospectively, and infrequency of the outcome of interest limited the number of covariates that could be included in models without over-categorising. Furthermore, vascular risk factors, especially smoking, were under-ascertained and treatment variables could be included only individually in models. Taken together, this leaves the potential for residual confounding.

Fors et al. 2016 investigated mortality in 255 individuals with schizophrenia. Strengths of the study included the thorough search for all people with schizophrenia in a catchment area, the use of comparisons with matched referents from the general population (without schizophrenia), and the thorough process of diagnosis and re-diagnosis by a psychiatrist. The study was performed in a middle-sized town and its surroundings, and may therefore be more representative of people with schizophrenia in Sweden than a study performed in a larger city. Dropouts were few: 1.5% over the follow-up period of 10 years. Weaknesses of the study included the lack of information in the medical data sources on confounders such as smoking habits and alcohol/drug abuse which may have contributed to the excess mortality, and lack of information on glucose tolerance or other signs of metabolic disturbance. Another problem was the lack of control for treatment outcome and changes in the healthcare organisation that could have impacted how healthcare was delivered.

Hjorthøj et al. 2015 investigated all-cause mortality and cause-specific mortality associated with substance use disorder, including alcohol and cannabis, in 41,470 individuals with schizophrenia and 11,739 individuals with bipolar disorder. The limitations include underestimation of the associations because substance use disorders are under-diagnosed. Furthermore, by limiting analyses to individuals with a diagnosed substance use disorder, the mortality effects of drinking or drug taking that do not meet the thresholds of substance use disorder were not assessed. Additionally, in the primary analyses, an immortality bias was potentially introduced because individuals in the exposed categories, who were diagnosed with substance use disorder, would have survived to the point of diagnosis, whereas individuals without diagnosed substance use disorder could have died at any point

during follow up. Therefore, the mortality rate in substance users will be under-estimated because they were observed for a longer period. The study classified people who had received prescription medicines used to treat alcohol and hard-drug use disorders as having the disorder in question but they may have been prescribed for a non-mental health reason.

Høye et al. 2011 investigated mortality ratios among 1,111 individuals with schizophrenia. There was minimal loss to follow-up as the Norwegian Cause of Death Register can be considered complete with regard to mortality. Being the only regional psychiatric hospital, all admitted residents should have been included, although non-admitted individuals with schizophrenia would not be included. This is particularly likely in the last part of the 27 year period because of more extensive out-patient care. Non-differential misclassification of diagnoses would have resulted in attenuation of the reported relationships.

Kim et al. 2017 investigated outcomes for all-cause and suicide mortality rates among 9,387 individuals with schizophrenia. The study used ICD-10 to classify medical illnesses; however, inaccuracies in diagnoses could have occurred as recordings made by different individuals. Information was not available on several key risk factors, including education level, smoking status, and alcohol consumption status. The date of death was only given up to the year and month in the data, therefore may have led to some inaccuracies in estimating date of death.

Lahti et al. 2012 investigated cardio- and cerebrovascular hospitalisations and mortality with reimbursement entitlements and use of drugs for treatment of hypertension, dyslipidaemia,

coronary heart disease and diabetes among 12,939 individuals with schizophrenia. The strengths of this study include the large, population-based sample and the long follow- up period. The diagnoses in the Hospital Discharge Register and the Causes of Death Register are well validated. However, although the overall sample was large, numbers of individuals in specific diagnostic categories were relatively low, reducing power. Furthermore, this study recorded the use of medication from prescriptions filled in out-patient care, not covering inpatient receipt, resulting in a potential classification bias.

Laursen et al. 2011 investigated hospital admission and mortality among 2,450,812 persons with schizophrenia and bipolar disorder. The study utilised diagnostic information in the Psychiatric Central Register for schizophrenia and bipolar disorder which is known to have a strong agreement between the psychiatric clinical diagnoses in the registers and research criteria diagnoses. In Denmark close to 100% of all diagnoses of schizophrenia and bipolar disorder are registered in the Psychiatric Central Register because the GPs in virtual 100% of the cases refer individuals with symptoms of schizophrenia or bipolar disorder to an in-patient or out-patients contact to a psychiatric hospital. Individuals in outpatient treatment tend to have many contacts, and the number of incorrectly assigned person-years would be limited. The study only included younger persons under the age of 52 because the Danish registries did not contain complete medical information on persons born before 1955. Similarly, Laursen et al. 2014 investigated mortality outcomes among 1,061,532 individuals with schizophrenia and bipolar disorder with cardiovascular drug use. The limitations include the fact that its association between psychiatric status and cardiac deaths could not be controlled for behavioural risk factors such as smoking, diet and physical inactivity. The

study measured exposure on first redemption of a medication and no estimations of compliance were made.

Medici et al 2015 investigated age- and gender standardised mortality incidence rates, mortality ratios and Kaplan–Meier survival estimates among 15,334 individuals with bipolar disorder. A strength was the fact that age at time of diagnosis was not subject to recall bias. However, mortality in prevalent cases cannot be fully assessed over relatively short periods of time, because of an increased risk of incidence-prevalence bias, where individuals with a long and relapsing clinical course dominate the picture; therefore, only incident cases were followed.

Ösby et al. 2016 investigated mortality adjusted for age, sex and time period among individuals with schizophrenia, bipolar or unipolar mood disorder from 1987–2010 (614,035 person-years). The strengths include the fact that Swedish Registers include all residents in the country and are considered comprehensive, with 99% of all somatic and psychiatric hospital discharges recorded. Swedish hospitals and government agencies are required by law to enter medical information to this register. All diagnoses in the national patient register and the Cause-of-Death Register were given by the physician treating the individual using an internationally accepted standard classification. Being register based, this study used information about clinical diagnoses from hospital admissions in Sweden, but the categorisation of patient groups was based on the most recent diagnoses given at the discharge. The diagnoses were based on in-patient diagnoses, which may have generated a

selective bias towards severely ill individuals; however, most individuals with schizophrenia and most with severe stages of bipolar and unipolar mood disorder are thought to be admitted to hospital in Sweden. Since medical care, including hospital care and prescribed medication, is heavily subsidised in Sweden, there is no bias by costs for hospital care leading to differences in health-seeking behaviour. The researchers did not have access to medical records or information on medical treatment, and lacked information on other factors affecting health, such as measures of socio- economic status and health behaviour.

Pan et al. 2016a investigated hospital treatment, mortality and healthcare costs among 7,987 individuals with bipolar disorder. The study strengths include whole country coverage, inclusion of bipolar disorder individuals diagnosed in all clinical settings, longitudinal follow-up for consecutive 3 years, examinations of socioeconomic status at both personal and household levels and provision of the rarely available data on the relationship between socioeconomic status, outcomes of hospital treatment and mortality. By including only newly diagnosed individuals, the study further minimised the bias in the baseline socio-economic status because of illness-related social decline. Similar strengths and limitations apply to the study reported by Pan et al. 2016b investigating mortality risk among 15,254 individuals with bipolar disorder, and that of Pan et al. 2017 investigating age- and gender-standardised mortality ratios among 272,617 individuals with schizophrenia and bipolar disorder.

Talaslahti et al. 2012 investigated mortality in 9,461 individuals with schizophrenia aged over 65 years. Its limitations included the fact that it was not possible to check the reliability of the schizophrenia diagnoses and no information was available on the living habits of the individuals or on the exact age at onset of the disorder. From a methodological point of view, because several statistical tests were performed, some possibility of chance findings (type I error) exists, but because of the small number of cases in some of the subgroup analyses, a possibility of type 2 error to detect small differences may also have occurred. Talaslahti et al. 2015 investigated mortality among 6,142 individuals with earlier onset schizophrenia with similar limitations, as well as the limited generalisability to late onset schizophrenia and other late life psychiatric conditions.

Tenback. et al. 2012 investigated all-cause mortality among 7,415 individuals with schizophrenia, strengths of which included the reliability of routinely collected data from a health insurer with no missing data. However, no data were available on risk factors such as length of illness, smoking, alcohol or substance abuse, socioeconomic status, poor nutrition, or diminished physical activity levels. In addition the follow-up period of 3 years is limited, most individuals in this database will have had longer illness duration, and individuals might have been switched from medication because of cardiovascular risk factor such as excessive weight gain.

Tiihonen et al. 2016 investigated all-cause and cause-specific mortality rates as a function of cumulative low, moderate, and high exposure to antipsychotics, antidepressants, and

benzodiazepines among 21,492 individuals with schizophrenia. The information was limited as the usage of open care health services could not be obtained from registers. The register data did not allow comparison of individuals with late onset schizophrenia and individuals with other old age psychiatric conditions.

Westman et al. 2013 investigated mortality rate ratios, excess mortality, cardiovascular disorder and specifically cerebrovascular disease, coronary heart disease, acute myocardial infarction, sudden cardiac deaths and hospital admission rate ratio among 17,101 individuals with bipolar disorder. The sample was limited to those with inpatient diagnoses, which may have generated a selective bias towards severely ill individuals. Authors additionally did not have access to medical records or information on medical treatment to ascertain additional variables.

2.4.4 Relative levels of all-cause mortality in SMI

The search identified 24 studies that investigated all-cause mortality as listed in Table 2.3. The comparators used in most studies were from the general population except in Schoepf et al., 2014 where age- and gender-matched hospital controls were included and Talaslahti et al., 2012 and 2015 which compared early onset with late onset schizophrenia population. Standardised mortality ratios (i.e. comparisons of mortality risks of people with SMI with the age specific general population), hazard ratios, odds ratios and relative risks were used.

Chang et al 2010 was the only study included in the review that included a mortality measure for individuals with schizoaffective disorder. Overall, the SMR for SMIs was 2.15 but when stratified by SMI diagnosis, schizoaffective disorder had the highest SMR of 2.52 and bipolar disorder had the lowest SMR of 1.95 (all SMRs statistically significant). As well as Chang et al 2010, the other study that investigated mortality in people with SMI as a whole found a raised SMR of 3.6 for all-cause mortality (Reininghaus et al., 2015).

Eleven further studies investigated risk of death among people with schizophrenia specifically. The studies reporting SMRs found that mortality was significantly high among people with schizophrenia, apart from the study reported by Dickerson et al., 2014 where SMRs were not significant. Among all other studies investigating schizophrenia mortality risk and reporting risk ratios or hazard ratios, deaths were significantly higher compared with general population. In the Hoang et al 2013 study, the individuals with schizophrenia were

younger than 75 years, and their SMR was higher than the studies that did not have an age cut-off. However, the Talaslahti et al 2015 study found that SMR was higher among individuals with late onset schizophrenia (SMR=5.02- 95% CI 4.61- 5.46) compared to earlier onset schizophrenia (SMR=2.93- 95% CI 2.83- 3.03).

Although reviewing change over time was not a primary aim, several studies included did investigate mortality measures at different time points. Pan 2017 found that the mortality gap for people with schizophrenia had decreased slightly, reporting an SMR of 3.40 (95% CI 3.30–3.50) for their 2003 cohort and 3.14 (3.06–3.23) for their 2008 cohort. On the other hand, Hoang et al 2011 reported that the mortality hazard ratio for schizophrenia in 1999 was 1.6 and in 2006 had increased to 2.2; the same study found hazard ratios for bipolar disorder in 1999 as 1.3 and in 2006 as 1.9 (Hoang et al 2011). In the Pan 2017 study, the mortality gap for bipolar disorder individuals remained relatively stable with only those aged 15–44 years having an SMR rising significantly from 7.04 (6.38–7.76) to 9.10 (8.44–9.79).

In relation to bipolar disorders, in all identified studies reporting mortality risk using SMRs, RRs and HRs, significantly high mortality risk was found among people with bipolar disorders compared with general population.

Table 2:3- All-cause mortality rates among individuals with SMI

Author	Rate/HR/OR/ SMR (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Chang et al., 2010	SMR=2.15 (1.95-2.36)2.25 (2.01-2.51, n = 322), 2.52 (1.83-3.39, n = 44), 1.95 (1.60-2.35, n = 108)	SMI (F20, 25, & 31 combined), Schizophrenia (F20), Schizoaffective disorders (F25), Bipolar affective disorder (F31)
Reininghaus et al., 2015	SMR=3.6, (2.6–4.9)	SMI
Callaghan et al., 2014	HR= 2.45 (2.41–2.48)	Schizophrenia
Castagnini et al., 2013	HR= 1.3 (1.0–1.6)	Schizophrenia
Dickerson et al., 2014	SMR=2.80 (0.89- 6.38).	Schizophrenia
Ernst et al., 2013	HR= 2.05 (2.01–2.09) HR=1.6 (1.5-1.8) in 1999	Schizophrenia
Hoang et al., 2011	HR=2.2 (2.0 to 2.4) in 2006	Schizophrenia
Hoang et al., 2013	SMRs= 4.23 (3.85–4.60)	Schizophrenia aged <75 years
Kredenster et al., 2014	RR= 1.70, P < .0001	Schizophrenia

Author	Rate/HR/OR/ SMR (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Olfsen et al., 2015	SMR= 3.7 (3.7-3.7)	Schizophrenia
	SMR for 2003 cohort= 3.40 (3.30– 3.50)	
Pan et al., 2017	SMR for 2008 cohort= 3.14 (3.06–3.23)	Schizophrenia
		Late onset schizophrenia in individuals aged
	SMR=5.02 (4.61–5.46) in late onset	65years or above vs earlier onset (age below
Talaslahti et al., 2015	SMR=2.93 (2.83–3.03) in earlier onset	60years)
Callaghan et al., 2014	SMR= 1.57 (1.53–1.62)	Bipolar disorder
	Women- HR=2.34 (2.16-2.53)	
Crump et al., 2013	Men- HR=2.03 (1.85- 2.23)	Bipolar disorder
	HR=1.3 (1.1 to 1.6) in 1999 and	
Hoang et al., 2011	HR=1.9(1.6 to 2.2) in 2006	Bipolar disorder
Hoang et al., 2013	Bipolar HR=2.60 (2.21–3.00)	SMI aged <75 years
		Bipolar patients who needed inpatient
Pan et al., 2016a	HR=1.74 (1.40–2.15)	treatment

Author	Rate/HR/OR/ SMR (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Pan et al., 2017	Aged 15–44years- SMR= 7.04(6.38–7.76) to 9.10(8.44–9.79). Aged 15–44 years- natural-cause-SMR= 5.65 (4.93–6.44) to 7.16 (6.46–7.91)	Bipolar disorder
Westman et al., 2013	MRR 2.40; 95% CI 2.33 to 2.47)	Bipolar disorder

2.4.5 Relative levels of cause-specific mortality in SMI

The review identified studies that calculated SMRs or HRs or mortality rate ratios as a measure for strength of association. Table 2.4 summarises the findings from the literature reporting on underlying cause of death. Overall there is a combination of natural and unnatural causes that are associated with mortality. Chronic physical conditions such as chronic obstructive pulmonary disease and malignant neoplasm were investigated as well as unnatural causes of death including suicide and accidental deaths. The evidence base was extensive for cardiovascular disease.

In the study by Reininghaus et al., 2015 which found an overall SMR of 3.6, the SMR for natural-cause mortality was 1.7 (95% CI 1.0–2.7) and unnatural-cause mortality was 13.3 (95% CI 8.7–20.4). Severe mental disorders are associated with elevated suicide rates. For example, the mortality rate due to suicide is estimated to be over 12 times greater among people with schizophrenia compared to the general population. In the publications summarised in table 2.3, a history of suicide attempts, depression, not taking medications as prescribed, and drug and alcohol misuse are risk factors for suicide among patients with schizophrenia and bipolar disorder.

Two studies investigated accidental deaths among individuals with schizophrenia and found significantly high SMRs compared with the general population. Olfsen et al, 2015 found SMR

of 3.2 for accidental deaths and Talaslahti et al, 2015 found a ten-fold difference in SMRs by onset groups: 33.74 in those with an onset at 60 years or later group compared to 8.93 in those with an onset before 60 years.

Six more studies investigated cancer mortality among patients with SMI and found SMRs were significantly higher than the general population. Das Munshi et al., 2017 found the cancer SMR to be 1.45 and Olfsen et al., 2015 reported an SMR of 2.4, while Kredenster et al., 2014 described a risk ratio of 1.45 for lung cancer mortality. Talaslahti et al., 2015 found an SMR of 1.75 among individuals with late onset schizophrenia aged 65 years or older while for individuals with earlier onset SMR was 5.05. and Castagnini et al., 2013 reported a hazard ratio of 1.9 for cancers. A further 11 studies investigated the SMRs for CVD, circulatory, stroke and cerebrovascular deaths in patients with SMI compared with the general population. In 8 studies out of 11, significantly raised SMRs were found compared with either the general population or a specific control group, ranging from an SMR of 3.6 in the study by Olfsen et al., 2015 in relation to cardiovascular disease, to a hazard ratio of 1.67 for stroke in schizophrenia reported by Lahti et al., 2012.

Five studies investigated SMRs for respiratory related deaths such as pneumonia, influenza, chronic obstructive pulmonary disease and general respiratory illness. All of these showed significantly high SMRs or risk ratios among patients with SMI compared with general or control populations. The risk ratio reported by Kredenster et al., 2014, was 2.0 for deaths in

respiratory illnesses, while the SMR was 9.9 for COPD deaths in the study reported by Olfsen et al., 2015.

Table 2:4- Summary of research findings evaluating cause-specific mortality in SMI

Author	Cause of death	Strength of the association with mortality (95%CI)	Study cohort (Cases Vs Controls (comparisons))
Olfsen et al., 2015	Accidental deaths	Accidental deaths SMR= 3.2 (3.1-3.2)	Schizophrenia
Talaslahti et al., 2015	Accidental deaths	SMR= 33.74 (20.88–51.58)	Late onset schizophrenia in individuals aged 65years or above vs earlier onset (age below 60years)
Castagnini et al., 2013	Bipolar- natural causes	Bipolar- natural causes HR= 1.5 (1.1–1.9)	Schizophrenia
Das- Munshi et al., 2017	Cancers	Cancers SMR=1·45 (1·32–1·60).	SMI
Kredenster et al., 2014	Cancers	28.6 vs. 27.3 per thousand, P = 0.42	Schizophrenia
Olfsen et al., 2015	Cancers	Among 6 selected cancers, lung cancer had the highest mortality rate (74.8 per 100,000 person-years) SMR= 2.4 (2.4-2.5).	Schizophrenia
Talaslahti et al., 2015	Cancers	SMR= 1.75	Late onset schizophrenia in individuals aged 65years or above vs earlier onset (age below 60years)
Castagnini et al., 2013	Cancers	Neoplastic diseases HR= 1.9 (1.1–3.2)	Schizophrenia
Kredenster et al., 2014	Lung Cancer mortality	RR = 1.45, P < 0.001	Schizophrenia

Author	Cause of death	Strength of the association with mortality (95%CI)	Study cohort (Cases Vs Controls (comparisons))
Das- Munshi et al., 2017	Cardiovascular disease	cardiovascular disease SMR=2·65 (2·45–2·86)	SMI
Fiedorowicz et al., 2009	Cardiovascular disease	Bipolar subtype no longer predicted cardiovascular mortality HR=1.85 (0.74–4.65), p=0.19	Bipolar disorder
Olfsen et al., 2015	Cardiovascular disease	Cardiovascular disease had the highest mortality rate (403.2 per 100,000 person-years) SMR=3.6 (3.5-3.6).	Schizophrenia
Brown et al., 2010	Cardiovascular disease	SMR at start=129 (27–377) to SMR at end=350 (186–598), P=0.053	Schizophrenia
Westman et al., 2013	Cerebrovascular disease	MRR=2.00 (1.81- 2.22)	Bipolar disorder
Fors et al., 2016	Circulatory diseases	RR= 2.2, p<0.001	Schizophrenia
Talaslahti et al., 2015	Circulatory diseases	SMR= 2.26	Late onset schizophrenia in individuals aged 65years or above vs earlier onset (age below 60years)
Kredenster et al., 2014	Circulatory illness	RR = 1.64, P < .0001	Schizophrenia
Lahti et al., 2012	Coronary heart disease	HR=2.92 (1.70–5.00)	Schizophrenia
Westman et al., 2013	Coronary heart disease	MRR= 1.95 (1.81- 2.09)	Bipolar disorder

Author	Cause of death	Strength of the association with mortality (95%CI)	Study cohort (Cases Vs Controls (comparisons))
Westman et al., 2013	Mortality from sudden Cardiac death, cardiac arrest/ventricular fibrillation	MRR=1.85 (1.42- 2.41)	Bipolar disorder
Lahti et al., 2012	Stroke	HR= 1.67 (0.53–5.26)	Schizophrenia
Kredenster et al., 2014	Injury	RR = 2.35, P < 0.0001	Schizophrenia
Olfsen et al., 2015	influenza and pneumonia	SMR= 7.0 (6.7-7.4)	Schizophrenia
Olfsen et al., 2015	Chronic obstructive pulmonary disease	SMR= 9.9 (9.6-10.2)	Schizophrenia
Das- Munshi et al., 2017	Respiratory diseases	SMR= 3·38 (3·04–3·74)	SMI
Talaslahti et al., 2015	Respiratory diseases	SMR= 6.18	Late onset schizophrenia in individuals aged 65years or above vs earlier onset (age below 60years)
Kredenster et al., 2014	Respiratory disease	RR = 2.00, P < .0001	Schizophrenia
Crump et al., 2013	Suicide	Female HR=10.37 (7.36-14.60) Male HR=8.09 (5.98-10.95)	Bipolar disorder
Høye et al., 2011	Suicide	SMRs=17.0 (12.8- 22.4)	Schizophrenia

Author	Cause of death	Strength of the association with mortality (95%CI)	Study cohort (Cases Vs Controls (comparisons))
Olfsen et al., 2015	Suicide	Suicide SMR=3.9 (3.8-4.1)	Schizophrenia
Westman et al., 2013	Suicide, homicides and accidents	MRR=9.66 (8.99 to 10.37)	Bipolar disorder
Das- Munshi et al., 2017	Suicides	Age-and-sex-standardised SMR=7·65 (6·43–9·04)	SMI
Kredenster et al., 2014	Suicides	RR = 8.67, P <0 .0001	Schizophrenia

2.4.6 Demographic factors as predictors of mortality in SMI

Age and gender have been extensively investigated and did not appear to present controversy in terms of outcomes. As listed in Table 2.5, the studies shortlisted for this review confirm that mortality in SMI is greater for males and a steady increase in mortality associated with increasing age as described in Koola et al., 2012, Fors et al., 2016 and Brown et al, 2010 except one study which showed risk ratio was 3.26 compared with females (Dickerson et al., 2014). Two studies investigated the association between ethnicity and mortality among patients with SMI and found no significant association with non-Caucasian compared with Caucasian ethnicity and differing directions of association.

Table 2:5- Summary of findings from literature examining demographic factors as a predictor of mortality

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Brown et al., 2010	Age	One year increase in age: OR= 1.13 (1.07-1.18)	Mortality among patients with SMI with co-morbid diabetes
Brown et al., 2012	Age	Age at onset-HR 1.08 (1.02-1.13)	Schizophrenia
Fors et al., 2016	Age	Age 45-64 years, HR=4.12, p<0.001, >65 years HR= 23.7, p<0.001 (reference group age 18-44 years)	Schizophrenia
Koola et al., 2012	Age	Older ages (35–55 years versus <35 years)- HR=2.3, p=0.01	Cannabis users vs non-users with psychotic disorders
Tenback et al., 2012	Age	Age per one year increase- HR= 1.06, (95% CI, 1.04–1.08)	Schizophrenia
Brown et al., 2010	Ethnicity	Non-Caucasian ethnicity vs Caucasian ethnicity- OR= 1.98 (0.97- 4.00)	Mortality among patients with SMI with co-morbid diabetes
Dickerson et al., 2014	Ethnicity	1.59 (0.64 – 4.82) for Caucasian ethnicity	Mortality among patients with SMI with co-morbid diabetes

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Dickerson et al., 2014	Gender	RR= **3.26 (1.45–7.98), for females**	Schizophrenia
Aagaard et al., 2016	Gender	OR= 1.5 (1.0-2.1) P= .031 for males	Schizophrenia and substance use
Brown et al., 2010	Gender	Male: 0R= 1.95 (95% CI 0.97- 3.90)	Mortality among patients with SMI with co-morbid diabetes
Davis et al., 2015	Gender	Male HR= 1.52 (1.29–1.78)	Schizophrenia
Koola et al., 2012	Gender	Males HR=2.2, p=0.02	Cannabis users with psychotic disorders

2.4.7 Lifestyle, socioeconomic/ social environmental factors and mortality in SMI

Table 2.6 lists the findings from literature examining lifestyle, socioeconomic/ social environmental factors as predictors of mortality in cohorts with SMI. Within the papers included in this review the following factors were assessed- body mass index, cigarette smoking, education, employment, family support, household income, occupational and recreational activities, social relationships, standard of living conditions and urbanisation.

One study examining the association between higher body mass index and mortality among individuals with schizophrenia found a 3% increase in hazard in each one-unit increase in BMI (Davis et al., 2015). The authors explain that the increase in BMI is partly due to unmonitored antipsychotic treatment, which can lead to weight gain. Significant weight gain is one of the main reasons individuals do not want to take prescription medication. Weight gain in this population also poses a significant risk of lipid abnormalities and cardiovascular complications.

Smoking was identified as adversely associated with mortality in six of these studies. All studies reporting this exposure found clear associations between risk of smoking and mortality among individuals with SMI. A risk ratio of 4.67 in Dickerson et al., 2014 among individuals with schizophrenia. SMRs of 2.45 and 1.57 was reported for individuals with schizophrenia and bipolar disorder, respectively, in Callaghan et al., 2014. The odds of a current smoker with SMI and co-morbid diabetes dying was found to be double that of a

person who is diabetic and is currently not smoking (Brown et al., 2010b). Even among all individuals with schizophrenia not known to have diabetes or other specific co-morbid conditions, the hazard ratio for mortality is 2.5 times higher for current smokers compared to individuals who not current smokers (Brown et al., 2012).

Years of education (Dickerson et al., 2014) and qualifications (Reininghaus et al., 2015) were both not statistically associated with mortality. Reininghaus et al., 2015 also reported that there was no statistically significant association with unemployment.

Bowersox et al., 2012 found that lack of re-engagement was associated with increased non-injury death (odds ratio 1.64), with increased cancer-related mortality (odds ratio 4.76) and with an average of 97.4 fewer days of life. Family involvement at first contact was found to be significantly protective with a risk ratio of 0.09 (Reininghaus et al., 2015). One study investigated household income and deaths among SMI patients and found that the SMR was 6.47 among females and 5.51 for males with SMI among low income families compared with those not from low income households (Pan YJ et al., 2016). Factors such as family support (Reininghaus et al., 2015) and social relationships (Hayes et al., 2012b) have not been found to have strong associations with mortality, although urbanisation was associated with higher mortality in patients with schizophrenia (Fors et al., 2016).

None of the studies in this review provided a comprehensive suite of factors. Findings on the different predictors have been mainly derived from studies using different methodologies and populations samples and should be interpreted with caveats.

Table 2:6- Summary of findings from literature examining lifestyle, socioeconomic/ social environmental factors as a predictor of mortality

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Davis et al., 2015	Body Mass Index	Increase of 1 kg/m2 Hazard ratio= 1.03 (1.01–1.04),	Schizophrenia
Brown et al., 2010b	Cigarette smoking	Current smoker vs not current smoker: OR= 2.06 (1.01-4.20)	Mortality among patients with SMI with co-morbid diabetes
Brown et al., 2012	Cigarette smoking	Current smoker vs not current smoker: HR 2.53 (1.01-6.34)	Schizophrenia
Callaghan et al., 2014	Cigarette smoking	Mortality estimates for conditions causally related to tobacco vs conditions not causally related to tobacco smoking: Schizophrenia SMR= 2.45 (2.41- 2.48), Bipolar SMR= 1.57 (1.53- 1.62)	General population vs SMI

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Davis et al., 2015	Cigarette smoking	Current smoker vs never or ex-smoker HR=1.98 (1.59–2.47)	Schizophrenia
Dickerson et al., 2014	Cigarette smoking	Cigarette smoker vs non-smoker: RR= 4.66, P =0 .0029	Schizophrenia
Dickerson et al., 2014	Education	Year of education: RR= 0.89 (0.78 – 1.03)	Schizophrenia
Reininghaus et al., 2015	Education	No qualifications: 1.32 (0.69–2.53)	SMI
Reininghaus et al., 2015	Employment	Unemployment: 0.70 (0.37–1.33)	SMI
Reininghaus et al., 2015	Family support	Family involved in first contact with health services: RR= 0.09, (95% CI 0.01–0.69)	First-episode psychosis
Pan Y-J 2016	Household income	SMR for low-income males=5.51 (3.21- 8.81) SMR for low-income females=6.47 (2.96, 12.28)	Individuals with bipolar who needed inpatient treatment

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Hayes et al., 2012b	Occupational and recreational activities	HR for Significant problem: 0.9 (0.6–1.3)	SMI patients aged 15 years and over
Hayes et al., 2012b	Social relationships	HR for Significant problem: 0.9 (0.6–1.3)	SMI patients aged 15 years and over
Hayes et al., 2012b	Standard of living condition	Significant problem: HR= 1.1(0.8–1.6)	SMI patients aged 15 years and over
Fors et al., 2016	Urbanisation	Living in the city than among those living in less urbanized areas- HR=2.46, p= 0.03.	Schizophrenia

2.4.8 Physical co-morbidities

Table 2.7 provides a summary of findings from studies which investigated physical co-morbidities as predictors of mortality in people with SMI. In general, having a comorbidity was associated with an increased risk of mortality.

Five of the studies reviewed had assessed circulatory conditions. Davis et al., 2015, reported that an increase of 10 mm Hg systolic blood pressure was associated with a decreased hazard ratio for mortality of 0.92 (0.88–0.96), although an increase of 5 mm Hg to diastolic blood pressure was associated with an increased hazard ratio of 1.07 (1.03–1.12). Lahti et al., 2012 found that hospital admission with Coronary heart disease was associated with higher mortality. We et al., 2015 found an increased risk of mortality with macrovascular complications for individuals who had diabetes and schizophrenia which was statistically significant but not statistically significant with microvascular complications. Schoef et al., 2014 reported that the odds of suffering from an ischaemic stroke was three times higher among individuals with schizophrenia.

Findings for respiratory conditions were obtained from two studies. Schoepf et al., 2014 found for individuals with schizophrenia that the odds of being diagnosed with bronchitis (2.6 times higher), chronic obstructive pulmonary disease (COPD) (2.8 times higher) and pneumonia (3.0 times higher). Dickerson et al., 2013 also found a raised risk ratio for respiratory conditions but this was not statistically significant.

Cancer findings included reports from Pan et al., 2016a for individuals with bipolar disorder a significantly raised hazard ratio of 2.14. Kredenster et al., 2014 did not report a significant association with risk ratio of 1.05 (p-value of 0.42).

Digestive conditions were reported by Schoef et al., 2014 with a 10-fold increase of having alcoholic liver disease associated with general hospital mortality among individuals with schizophrenia. Other conditions assessed included diabetes as well as dermatological, genitourinary, haematological, immunological, musculoskeletal, neurological and renal conditions. Five studies included diabetes with statistically significant increased hazards ratios of 1.61 (Pan e al., 2016a) and 1.33 (Davis et al., 2014). The findings for dermatological conditions was not statistically significant (Dickerson et al., 2014). Risks were higher for genitourinary conditions (p-value 0.002) and haematological conditions (p-value 0.085). Schoef et al., 2014 also found increased odds of iron deficiency anaemia, a haematological condition ($p<0.05$). One study found evidence for raised risk of musculoskeletal conditions with risk ratios of 3.24 95% CI 1.18- 7.71) (Dickerson et al., 2013). The same study found no association with neurological conditions 2.04 (95%CI 0.83- 4.59). Renal disease has been significantly associated with schizophrenia- four time more renal failure (Schoef et al., 2014) and bipolar disorder- increased deaths from renal disease (Pan et al., 2016b).

Table 2:7- Summary of findings on physical co-morbidities as predictors of mortality in SMI

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Schoepf et al., 2014	Alcoholic liver disease	OR = 10.3, (3.3–32.1), p< 0.001	Schizophrenia general hospital mortality
Schoepf et al., 2014	bronchitis	OR = 2.6, (1.2–5.3), p<0.05	Schizophrenia
Pan et al., 2016a	cancer	HR=2.14 (1.74–2.63)	Bipolar disorder
Davis et al., 2015	CHD: cardiovascular disease	SBP (increase of 10 mm Hg) HR= 0.92 (0.88–0.96), DBP (increase of 5 mm Hg) HR= 1.07 (1.03–1.12)	Schizophrenia
Lahti et al., 2012	CHD: coronary heart disease hospitalisation	HR= 1.65 (1.03–2.57)	Schizophrenia
Schoepf et al., 2014	COPD	OR = 2.8, (1.7–4.7)	Schizophrenia

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Wu et al., 2015	CVD: complications	Macrovascular complications- HR- 1.49, (1.32–1.68), Microvascular complications- HR- 1.05, (0.91–1.21)	Diabetes & schizophrenia
Kredentser et al., 2014	Cancer	RR= 1.05 (p=0.4247)	
Dickerson et al., 2014	Dermatologic	RR= 0.53 (0.03–2.49	Schizophrenia
Pan et al., 2016a	Diabetes mellitus	HR=1.61 (1.38–1.89)	Bipolar disorder
Brown et al., 2010	Diabetes: Duration of diabetes	0R= 1.05 (1.01-1.10)	SMI with co-morbid diabetes
Brown et al., 2010	Diabetes: Hospitalised for diabetes-related condition	OR: 5.44 (1.94- 15.2)	SMI with co-morbid diabetes
Schoepf et al., 2014	Diabetes: Type 1 diabetes mellitus	OR = 3.8, (1.5–10.1)	Schizophrenia
Davis et al., 2014	Diabetes: Type 2 diabetes mellitus	HR= 1.33 (1.12–1.57)	Schizophrenia

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Dickerson et al., 2014	Genitourinary disease	RR = 4.35, P = 0.002	Schizophrenia
Dickerson et al., 2014	Haematological	RR = 3.46, P=0.085	Schizophrenia
Dickerson et al., 2014	immunologic disease	RR = 3.14, P = 0.044	Schizophrenia
Dickerson et al., 2014	Infections: level of antibodies to Epstein–Barr virus and to Herpes Simplex virus type 1	Epstein–Barr virus RR = 1.22, P = 0.0041 Herpes Simplex virus type 1 RR = 1.19, P = 0.030;	Schizophrenia
Schoepf et al., 2014	Infections: pneumonia	OR = 3.0, 1.6–5.6, p<0.001	Schizophrenia
Schoepf et al., 2014	iron-deficiency anaemia	OR = 2.8, (1.1–7.3), p<0.05	Schizophrenia
Schoepf et al., 2014	ischaemic stroke	OR = 3.3 (1.2–8.8), p<0.05	Schizophrenia
Pan et al., 2016a	Number of outpatient clinic visits within the first year of bipolar diagnosis	HR=1.74, (1.40- 2.15).	Bipolar disorder

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Dickerson et al., 2013	Musculoskeletal	RR= 3.24 (1.18–7.71)	Schizophrenia
Dickerson et al., 2013	Neurological	RR= 2.04 (0.83–4.59)	Schizophrenia
Schoepf et al., 2014	Parkinsonism	OR= 7.5 (3.2–17.4), p<0.001	Schizophrenia
Pan et al., 2016b	Renal disease	HR=1.65, (1.36–2.00)	Bipolar disorder
Schoepf et al., 2014	Renal disease: Non-specific renal failure	OR=4.7 (2.0–11.0), p<0.001	Schizophrenia
Dickerson et al., 2013	Respiratory	RR= 1.72 (0.63–4.08)	Schizophrenia

2.4.9 Psychiatric factors including co-morbidities

Factors that can be best grouped as 'psychiatric' have been summarised in Table 2.8. these include presence of lifetime manic spectrum episodes, cognitive function, depression and substance use including alcohol). Eleven studies included in this review had findings of relevance to substance misuse, alcohol, cannabis abuse. These studies indicate that, in general, a dual diagnosis of substance use disorder and SMI significantly increases mortality. Hjorthøj et al., 2015 study found that the standard mortality ratios were higher for individuals with schizophrenia (SMR= 8.46) was higher than that for individuals with bipolar disorder (SMR= 6.47). In other studies that have included hazard ratios for substance use disorders, the figures ranged from 1.74 (Pan et al., 2016b) to 6.81 (Dutta et al., 2007). The "non-substance use but psychiatric" factors investigated and included here are cognitive score, community treatment order, depression, remission, lifetime manic spectrum episodes, psychiatric hospitalisation as well as a comparison within SMI between bipolar and schizophrenia. One study of this review reported a protective effect from cognitive function among individuals with (Dickerson et al., 2014). Co-morbid depression is a significant risk factor for mortality among individuals with bipolar disorder with hazard ratio of 1.16 (p=0.02) (Fiedorowicz et al., 2014).

Table 2:8- Summary of findings from literature examining psychiatric factors including co-morbidities as a predictor of mortality

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Ramsey et al., 2013	Lifetime manic spectrum episodes	OR=1.42 (1.18- 1.70)	Bipolar with manic spectrum episodes
Dickerson et al., 2014	Cognitive function	RR = 0.96, P =0 .013	Schizophrenia
Fiedorowicz et al., 2014	Depression	HR= 1.16 (1.02-1.33), p = 0.02	Bipolar I disorder
Aagaard et al., 2016	Drug and/or alcohol use- Substance use disorder	RR= 0.93 (0.42–2.08)	Schizophrenia
Dickerson et al., 2014	Drug/alcohol abuse	RR= 6.61 (1.33–32.77)	SMI
Reininghaus et al., 2015	Illicit drug use- Substance use disorder	RR= 2.31, (1.06–5.03)	SMI

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Dutta et al., 2007	Alcohol abuse- Substance use disorder	HR 6·81, (1·69–27·36), p=0·007	Bipolar I disorder
Aagaard et al., 2016	Substance use disorder	OR= 1.8 (1.5- 2.2).	Schizophrenia and substance use
Hjorthøj et al., 2015	Substance use disorder	SMR = 8·46 (8·14–8·79)	Schizophrenia
Hjorthøj et al., 2015	Substance use disorder	SMR= 6·47 (5·87–7·06)	Bipolar disorder
Hjorthøj et al., 2015	Substance use disorder- alcohol, cannabis, hard drugs, combined	Alcohol- HR 1·52, (1·40–1·65); cannabis, 1·24, (1·04–1·48); hard drugs HR= 1·78, (1·56–2·04),	Schizophrenia
Hjorthøj et al., 2015	Substance use disorder- alcohol & hard drugs	HR= 1·52 (1·27–1·81) & HR= 1·89 (1·34–2·66)	Bipolar disorder
Pan et al., 2016b	Substance use disorder	HR=1.74, (1.37–2.21)	Bipolar disorder

2.4.10 Medication use

Six studies included in this review have investigated the effect of medication on deaths among individuals with SMI. The medications include cardiovascular medication such as lipid lowering agents, antihypertensive or diabetes medication, and psychiatric medication such as antidepressants, antipsychotics and specific medications including benzodiazepines, clozapine, lithium and other mood stabilizers. Table 2.9 provides a summary of findings from literature examining medication use as a predictor of mortality. The review identified two studies that specifically investigated medication for physical illness- lipid lowering medication and antihypertensive medication. Six other studies listed in the table presented outcomes for antidepressants, antipsychotic medication including clozapine, benzodiazepines, lithium and mood stabilizers. Lipid lowering medications were found to be associated with lower mortality risk in SMI (Lahti et al., 2012). In terms of mood stabilizers, Morgan et al., 2003 reported that lithium users with SMI had better survival while Tenback et al., 2012 found an association between mood stabilizers and increased risk of death. Tiihonen et al., 2016 found that moderate antidepressant exposure was associated with lower mortality. Similarly, clozapine was associated with lower mortality in Hayes et al., 2015. However, Tenback et al., 2012 found that receipt of first-generation antipsychotics was associated with higher mortality, and Tiihonen et al., 2016, found that treatment with benzodiazepine was associated with higher mortality among individuals with schizophrenia.

Table 2:9- Summary of findings from literature examining medication use as a predictor of mortality

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Lahti et al., 2012	Lipid lowering medication use	OR= 0.47, (0.27–0.80)	Schizophrenia
Lahti et al., 2012	Antihypertensive	HR= 0.37 (0.22–0.61).	Schizophrenia
Lahti et al., 2012	Diabetes medication	HR= 1.52 (0.93–2.51)	Schizophrenia
Tiihonen et al., 2016	Antidepressant- moderate exposure & high exposure	HR=0.85, (0.73–0.98) & HR=0.71, (0.59–0.86).	Schizophrenia
Tenback et al., 2012	Antipsychotics- first generation	HR= 2.36, (1.38–4.04)	Schizophrenia
Tiihonen et al., 2016	Antipsychotics moderate exposure & high exposure	HR=0.59 (0.49–0.70) &	Schizophrenia with varying levels of antipsychotic exposure

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
		HR=0.75 (0.63–0.89).	
Tiihonen et al., 2016	Benzodiazepines	HR= 1.74 (1.50–2.03)	Schizophrenia
Hayes et al., 2015	Clozapine	HR= 0.4 (0.2-0.7), p =0 .001	Schizophrenia
Morgan et al., 2003	Lithium	Univariate RR=0.39 (0.17-0.93), p=0.03, HR=0.37 (0.16-0.88), p=0.02 .	Schizophrenia, including both inpatients and those living in the community
Tenback et al., 2012	Mood stabilizers	HR= 8.42 (3.06–24.07)	Schizophrenia

2.4.11 Symptom profiles and other features of the mental disorder

Few studies were found to have investigated associations between symptoms and mortality in SMI. Table 2.10 provides a summary of findings from literature examining presenting broader features of mental disorders as predictors of mortality. Impairment of activities of daily living, depressed mood, hallucinations and delusions and overactive–aggressive behaviour has been investigated through extracting data from brief Health of the Nation Outcome Scales (HoNOS) on individual patient records. There have also been investigations of manic/hypomanic symptomatology, and deterioration from pre-morbid level of functioning up to a year after onset in the studies included in this review. Hayes et al., 2012b, reported that significant ADL impairment was associated with higher mortality in SMI including schizophrenia specifically. Another study by the same group (Hayes et al., 2012a) found that hallucinations and delusions and depressed mood had no effect on mortality in SMI. One study found that deterioration from pre-morbid level of functioning up to a year after the onset of bipolar conditions was significantly associated with higher mortality (Dutta et al., 2007). Fiedorowicz et al., 2009, investigated the association between manic/hypomanic symptomatology measured by affective symptom burden (the proportion of time with clinically significant manic/hypomanic or depressive symptoms) and cardiovascular mortality and found that there is high mortality due to CVD among bipolar patients (HR=1.30, 95% CI1.09–1.55).

Table 2:10- Summary of findings from literature examining presenting symptoms as a predictor of mortality

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Hayes et al., 2012b	ADL significant impairment	HR=1.9 (1.3–2.8) p=0.001	All SMI cases (schizophrenia, schizoaffective and bipolar disorder) aged≥15 years
Hayes et al., 2012b	ADL significant impairment	HR= 1.8 (1.2– 2.9)	Schizophrenia
Hayes et al., 2012b	ADL significant impairment	HR= 3.2 (1.1– 9.5)	Schizoaffective disorder
Hayes et al., 2012b	ADL significant impairment	HR= 2.0 (0.8– 4.9)	Bipolar affective disorder
Hayes et al., 2012a	Depressed mood- sub-clinical & severe	HR= 1.2 (0.9–1.7), p=0.174 & HR= 0.9 (0.6 –1.3), p=0.534	SMI cases (schizophrenia, schizoaffective and bipolar disorder) aged≥15 years
Dutta et al., 2007	deterioration from pre-morbid level of functioning up to a year after onset	HR= 5·20 (1·24–21·89), p=0·024	Bipolar I disorder

Author	Risk factor	Strength of the association with mortality (95% CI)	Study cohort (Cases Vs Controls (comparisons))
Hayes et al., 2012a	Hallucinations and delusions- subclinical & severe	HR=1.1 (0.7–1.6), p=0.773& HR= 1.0 (0.7–1.6), p=0.990	SMI cases (schizophrenia, schizoaffective and bipolar disorder) aged≥15 years
Fiedorowicz et al., 2009	Manic/hypomanic symptomatology	HR=1.30 (1.09–1.55), p<0.01	Bipolar disorder
Hayes et al., 2012a	Overactive–aggressive behaviour- subclinical & severe	HR= 0.8 (0.6–1.2), p=0.333& HR= 1.1 (0.8- 1.6), p=0.604	SMI cases (schizophrenia, schizoaffective and bipolar disorder) aged≥15 years

2.5 Strengths and limitations of the review

Various large data sources including case registers from the United Kingdom and Scandinavia were used in the studies included. Other studies were mainly carried out in developed 'Western' countries such as the USA, Finland, Denmark and Sweden; however, the review also included studies from South Korea and Taiwan. Sample sizes of the individual studies varied from less than 100 to over 1,000,000. The studies with larger sample size tended to obtain data from civil registers, a health register or an insurance database. The SMI population may, however, not be homogenous. For example, some studies included had data from community samples, inpatients only as well as utilised a broad range of data sources including Veteran registers, general hospital records and psychiatric hospital records leading to heterogeneity.

In terms of defining SMI, most studies used ICD criteria. Some of the older publications used ICD-8 or ICD-9 classification systems while recent cohorts have used ICD-10 classifications; few have used DSM-III-R or DSM-IV criteria. In terms of SMI cohort identification those studies used ICD have used schizophrenia (ICD-10 code: F20.x), schizoaffective disorder (F25.x) or bipolar affective disorder (F31.x). Some studies included F21.x and F23.x and I did not exclude them from the review as long as F20.x, F25.x or F31.x was included. Regarding all-cause mortality and cause specific mortalities, the search criteria only obtained the studies with predictors of SMI deaths and I did not carry out an

extensive search to review SMRs of all studies published to find out all-cause mortality or cause specific mortality SMRs. In terms of predictors, certain outcomes were extensively studied including demographics such as age and gender, alcohol and substance misuse, physical comorbidities such as diabetes.

Reflecting on the studies based on clinical records, as described above, the sample size varied considerably between studies. The logistical demands of ascertaining epidemiologically complete and accurate datasets has been challenging. Those studies with smaller sample size had reduced statistical power but were generally able to use supplementary data collection options to improve completeness. However, for the larger longitudinal studies, without technologically advanced mechanisms to extract the required data from structured fields or free text, utility has failed to reach full potential. The individual study designs introduced specific errors. In particular, for the ones using clinical records, mortality risk estimates are sensitive to the age composition of the study population and the length of follow-up.

Reflecting on studies based on registers, they have tended to be large, population-based and can include long follow-up. Completeness of data and consistency in information quality available further add to strengths. Several researchers have highlighted the risk of confounding or selection bias due to the non-randomised study design. In the absence of rigorous experimental methodology, it is not possible to be certain that the findings of an observational study are as indicative of a causal relationship as would be the case for

findings from a well-designed randomised controlled trial. Therefore, the majority of the findings described here have to be viewed as hypothesis-generating and different study designs may be needed to accurately establish causality.

2.6 Conclusion

This review of literature revealed several important issues. To begin with, there is consistent evidence that all-cause mortality for individuals with schizophrenia and bipolar disorder is higher than that observed in the general population. While it is true that the research field may need to move towards designing and evaluating interventions to improve this situation, it is also evident that the distribution of mortality risk in SMI remains poorly understood. Although there has been considerable previous research that has investigated factors that characterise and predict mortality, previous studies have been limited by investigating relatively few factors in the same study, mostly due to the unavailability of diverse contextual information. This means that findings on predictors of mortality in SMI have been mainly derived from studies using different methodologies and samples. In addition, studies have often been limited in the number of potential confounders investigated, which raises queries about unobserved confounding. Furthermore, although there have been some large cohort studies, generalisability of findings has been difficult to conclude beyond the populations that were examined, due samples focusing on specific population groups. This is particularly true if only inpatients are assessed or if insurance claims data is utilised. Therefore, it remains unclear whether key patient factors such as gender, ethnicity, social deprivation and clinical symptoms, predict mortality in other patient populations.

The evidence identified and reviewed does not sufficiently clarify whether there are any clinical subgroups of individuals with SMI who are at higher risk of mortality, or whether the raised risk applies equally to all individuals with SMI. There is a need for large cohort studies that are designed with minimal introductions of chance, bias, confounding and reverse causality and are generalisable to the population seen by secondary mental health care, which in the United Kingdom provides the majority of mental health care to individuals with SMI (NICE 2014). The studies need to be able to measure simultaneously a diverse number of possible predictors and outcomes in the same sample of patients as well as a range of possible confounders, thus improving the robustness of findings. Large studies are also needed to provide sufficient power to detect an association with rare outcomes such as mortality and to adjust for multiple confounders simultaneously in the analysis.

3 Rationale for the thesis

The South London and Maudsley (SLaM) hospital has used electronic records comprehensively since 2006 across all its services. The Clinical Record Interactive Search (CRIS) is a system developed to enable researchers to search and retrieve anonymised electronic health records efficiently. CRIS currently contains records for over 400,000 service users and allows the generation of bespoke research databases drawn from a pseudonymised copy of SLaM's EHR, including over 3.5 million de-identified documents in common word processor formats. Since its development, CRIS data have been substantially enhanced through external linkages and natural language processing. This provides unique opportunity to marry the benefits of clinical records based studies with those of the registry type studies, typically noted from the Scandinavian countries from the review in chapter 2, which are generally based on diagnoses and data derived from financial transactions in the health service, welfare services and other domains of public sector. In essence, the potential for a unique contribution from CRIS is embedded in the novelty of the methods where very large data can be automatically collected from medical records including correspondence between psychiatric services and general practitioners on to quantitative databases. However, CRIS has the possibility to go beyond diagnoses, and explore for example individual symptoms or traits, as described in later chapters. The more specific rationale for individual studies in this thesis is as follows:

1. As a first consideration, although it is recognised that natural and unnatural causes of death are responsible for premature mortality, there has been only limited evidence on the contribution of different more specific causes. In particular, while the life-expectancy gap has been described and widely cited (Chang CK, et al 2011), it is not clear how different causes of death account for this. Life expectancy is a product both of risk factors for mortality and of the age groups at which these deaths manifest. Thus, a disorder causing deaths in young adulthood will be responsible for many more years of life lost than a disorder of similar prevalence which primarily affects people in old age. In reviewing causes of death in SMI, I was not able to identify any previous piece of research which had evaluated this.

2. Medical comorbidity is an important factor in accounting for the lost life expectancy in SMI (Bahorik AL, et al 2017). Hospitalised medical conditions are likely to be particularly important because of their severity and sometimes life-threatening nature. However, in the course of my appraisal of the literature, I was not able to find any comprehensive description of acute care hospitalisation (i.e. inpatient episodes excluding mental healthcare) in people with SMI. While there has been more targeted research on individual comorbid conditions- some examples include diabetes (Brown C et al, 2010) and other cardiovascular risk factors (Ösby et al. 2016), I felt that it was important to generate at least some information on the most common reasons for hospitalisation across all physical healthcare diagnostic groups (as well as their relative frequency in people with SMI compared to the general population) in order to help inform priorities for intervention.

3. Finally, as concluded from the review described in Chapter 2, evidence remains limited on characteristics of people with SMI who are most at risk of adverse health outcomes – namely mortality. Most of the research in this area has focused on other health conditions, lifestyle factors such as smoking and substance use, and medications, as well as differences between diagnostic groups. This has largely been driven by data which happen to be available in the large administrative or healthcare databases used for mortality research in SMI. In this respect, I sought to take advantage of ongoing work in the mental healthcare database to develop more detailed clinical profiles using natural language processing. More specifically, I sought to investigate whether particular symptom profiles in SMI were associated with mortality and acute care hospitalisation, using mental healthcare hospitalisation as a readily available comparator outcome.

4 Objectives

The overarching aim of this thesis was to use a novel approach with detailed electronic clinical health records in a "proof of concept" for the use of natural language programming for data gathering in the investigation of adverse outcomes among individuals with severe mental illness. I used recent mortality and morbidity data for individuals with SMI and sought to answer the following questions:

- What are the relative contributions of major disease groups to the gap in life expectancy between individuals with SMI and the general population (chapter 6)?
- What are the most common reasons for admission to non-psychiatric hospitals by individuals with SMI and what are the relative frequencies of these admissions compared to the general population (chapter 7)?
- As a proof of principle for ascertaining meaningful symptom profiles from routine mental health record text fields, what is the prospective association between the number of recorded negative symptoms and mental healthcare outcomes (admission, duration of admission, and readmission) among individuals with schizophrenia (chapter 8)?
- What is the association with mortality and hospitalisation for each of six symptom dimensions (positive, negative, manic, disorganisation, catatonic and depressive) extracted from the clinical records of individuals with SMI (chapter 9)?

5 Methods

5.1 Chapter outline

In this chapter, I describe the study setting, main data source, main data extraction methods, study variables extracted from data sources and brief overview of statistical methods. The brief overview of the main statistical analyses will be described in more detail in individual analyses chapters. Data used in the analyses described in Chapters 6, 7, 8 and 9 were derived from the Clinical Record Interactive Search (CRIS) data resource, developed within the NIHR Mental Health Biomedical Research Centre and Dementia Unit (BRC/U), which allows researchers to search and retrieve de-identified clinical data sourced from the South London and Maudsley electronic health records. In this chapter, I have included a description of the operational model of the CRIS interactive search tool and natural language processing developments that have been instrumental to data described in this thesis. CRIS is a dynamic database, which updates against source files every 24 hours. At the time of writing, it contains more than 280,000 cases.

Some contents of this chapter have contributed to the following publication:

Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, Roberts A, Dobson RJ, Stewart R. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ Open. 2017 Jan 17;7(1):e012012. doi: 10.1136/bmjopen-2016-012012.

5.2 Setting

The South London and Maudsley NHS Foundation Trust (SLaM) is a mental health care Trust in London, which provides near-monopoly secondary mental healthcare for a geographic catchment of four London boroughs – Lambeth, Southwark, Lewisham and Croydon and a source population of over 1.2 million people (Stewart et al, 2009). It is believed to be the largest unit mental health care provider in Europe and provides a broad range of mental health services including inpatient, community, general hospital liaison and forensic services.

5.3 Data sources

Electronic Health Records and the Patient Journey System

All clinical notes at SLaM have been in the form of Electronic Health Records (EHRs) since 2006, using the 'Patient Journey System' (PJS) which was developed to integrate EHRs across all services in SLaM. Some SLaM services used EHRs before that time and legacy data were imported at roll-out. The Patient Journey System as the name suggests has enabled a patient's journey within the mental health service to be recorded and shared across multidisciplinary teams. PJS contains a comprehensive record of patient's demographics, contact information, referrals, transfers, detailed clinical assessments, care plans, medication, clinical activity and reviews in a combination of structured fields (such as dates and drop down menus) and unstructured free-text (e.g. case notes and correspondence) (Stewart et al. 2009). EHR-derived databases have the potential for revolutionising clinical research opportunities as they contain unprecedented large volumes of diverse real-world data and enable research to be carried out not only with source records but also with linkages to other data sources as described below.

Clinical Record Interactive Search

From 2006 onwards, electronic clinical records have been used comprehensively across all SLaM services, and in 2008, the Clinical Record Interactive Search (CRIS) system supported by SLaM's NIHR Biomedical Research Centre (BRC) for Mental Health was developed to enable researchers to search and retrieve anonymised electronic health records efficiently.

CRIS currently contains records for over 280,000 service users and allows the generation of bespoke research databases drawn from a pseudonymised copy of SLaM's EHR, including over 3.5 million de-identified documents in common word processor formats. Since its development, CRIS data have been substantially enhanced through external linkages and natural language processing (Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, Fernandes A, Hayes RD, Henderson M, Jackson R, Jewell A, Kadra G, Little R, Pritchard M, Shetty H, Tulloch A 2016), as will be described.

There is a security model for CRIS which has been developed and continues to be managed by an Oversight Committee. The committee includes membership from two mental health service users, a SLaM Research and Development Office representative, a Caldicott Guardian representative for the Trust, a Child and Adolescent Mental Health representative, the CRIS academic lead and CRIS manager. All potential CRIS projects are considered at this meeting following online application. Scientific robustness and patient confidentiality are of critical importance for a project to be approved and for the researcher to access CRIS within the SLaM firewall (Fernandes et al. 2013).

CRIS de-identifies PJS records not only by removing or condensing identifier fields (e.g. date of birth truncated to year and month), but also by identifying, marking and masking patient identifiers in text with ZZZZZZ (and carer identifiers with QQQQQQ), thus permitting the use of clinical data for research purposes within an overarching research governance framework. The algorithm used for de-identification has been evaluated and has achieved

98.8% masking of personal identifiers and 97.6% of all personal identifiers in text fields (Fernandes et al. 2013).

Data are therefore available in CRIS from both structured and unstructured fields from PJS. Structured data include information that is recorded as dates and in drop-down menu fields whilst unstructured data refer to free-text fields. Incorporating a daily update against its source files, CRIS provides up to date longitudinal information at individual patient level (Perera et al. 2016). The longitudinal nature of the data allows for factors to be measured at multiple points and for temporality to be determined.

5.4 Data extraction from CRIS

5.4.1 Natural language Processing

Natural language processing (NLP) techniques can be used for extracting knowledge from unstructured text data and have been developed in CRIS because most information of value in mental healthcare tends to be recorded as text. Of the many NLP techniques, information extraction has been used to convert unstructured text into structured data tables that can be used in database assembly – for example, ascertaining mentions of medications in text so that samples can be extracted on the basis of medication mentions and/or variables generated within samples to this effect. The technique can reduce time required by researchers to unlock information held in clinical notes, and represents the only way to achieve this when sample sizes move beyond what can be manually coded. In order to implement NLP processes within the local system (i.e. CRIS) a core NLP infrastructure is required. General Architecture for Text Engineering (GATE) provides this infrastructure and enables a text-mining capability within CRIS. GATE is a widely used suite of open source software for text engineering that includes a workbench for developing applications, tools for distributing those applications on different computer hardware architectures, a quality assurance suite and facilities. The GATE suite also includes tools for text-mining workflow, distributed processing and visualisation. A variety of text processing tools and document formats may be plugged into this architecture, with individual tools being chained together into processing 'pipelines', and documents processed in series through these pipelines. Two distinct shallow language processing methodologies have been adopted for CRIS development, in collaboration with University of Sheffield Department of Computer Science.

The first may be described as rule-based pattern matching of key concepts. Sentences are first processed to find and create annotations based on simple surface linguistic information (such as words, sentences, etc.). This step is then followed by the process of finding concept-specific keywords, which are used to recognise likely sentences of importance to the task followed by a set of patterns specific to the text-mining task are run over the previously generated annotations to create a final annotation containing all the information required in a readily extractable format. Given the complex nature of the topic in hand, the approach requires text-processing algorithms to be built manually by GATE users with language engineering skills, working closely with clinical and research staff.

More recently, local developments took place at the SLaM BRC with the development of TextHunter which is an NLP information extraction suite implemented jointly by SLaM and the Institute of Psychiatry, Psychology & Neuroscience at King's College London. TextHunter has been described in technical detail (Jackson et al, 2014), but in summary its principal purpose is to provide a user-friendly interface to accomplish three tasks required to extract concepts from free text: first, to find instances of a concept in a database of documents; second, to provide an efficient interface to allow human annotators to label a portion of the instances in order to develop a gold standard and training set; and third, to attempt to construct an appropriate statistical language model of the concept in order to generate structured data more widely. The TextHunter program typically provided the opportunity to annotate constructs to the following categories: (1) positive (coding text implying that the construct is present); (2) negative (coding text stating that the construct is absent); and (3) irrelevant text. TextHunter functionality also includes platforms for inter-annotator

agreement testing, and the creation of gold standard and test annotation sets. Whether rules-based or machine-learning approaches are used, separate training and test data sets are constructed. Symptom definitions that were developed from initial annotation guidelines are shown in the appendix 1.

Two key metrics are traditionally used in NLP for evaluating the information extraction application performance in the test data sets, at the level of the individual text annotation. *Precision*, broadly equivalent to positive predictive value in psychometrics, is the proportion of application 'hits' which are found to identify the genuine construct. *Recall* is the proportion of instances of the genuine construct which are identified by the application (i.e. equivalent to sensitivity). For precision and recall testing, a baseline level of performance of the application and later stage level measuring performance after post-processing have been helpful to monitor because advancements inevitably lead to a trade-off between the two. However, in the type of NLP applied in clinical records, suboptimal recall at an instance-level can sometimes be accepted as a price to pay for maximising precision because the longitudinal nature of EHR data means that there are generally multiple opportunities for an NLP application to capture a piece of information (e.g. if an NLP algorithm missed a symptom being stated as present in one sentence, it might well be picked up later in the document or record). Clearly this depends to an extent on the use to which an NLP algorithm is to be put.

The most common approaches in TextHunter and GATE applied to CRIS have used a "bag-of-words" feature which includes identifying keywords, surrounding word tokens and part-of-speech tags in conjunction with knowledge engineering features to build a Support Vector Machine (SVM) sentence classifier. SVM is a machine learning technique where the intention is to represent instances of text as vectors in high dimensional space. With a training set of instances labelled as indicative of a desired class, the SVM implementation generates a hyperplane which can in turn be used to classify unseen instances pertaining to the described class in the training set. SVMs have been used in CRIS where the assertion to be extracted is simple and restricted to a concise set of clinical language. In these cases, the specific NLP task is defined as a sentence classification problem, with a classifiable instance as a sentence containing (for example) a symptom keyword, or the general constructs of a symptom domain. In addition to the keywords, clinically relevant modifier terms are also defined for some concepts, to produce sub-classifications of symptoms where appropriate (Table 5-1). If a modifier term is detected within 8 words of a keyword, the modifier is deemed to be a possible relation. Further specifications have been applied so that modifiers can be 'mandatory' (meaning a modifier is required to be present for the definition of an instance to be met), or 'optional' (meaning only the keyword needs to be present for an instance definition to be met).

Considering NLP used to define symptoms (a key consideration for this thesis) and regarding potential biases that might result from missing synonyms outside of the selected keywords, these have not been considered to be a significant problem because the intended end-point was to identify clinician-assigned constructs, rather than attempt to classify descriptions of

experiences. In addition, keywords that are standard methods of symptom documentation by clinical staff, to avoid uncertainty over the authors' intent or meaning were included.

Table 5:1- Examples of natural language processing constructs applied in CRIS (Abstracted from (Jackson et al. 2017))

Modification type	Keyword	Modifier	Example of a 'positive' annotation
Mandatory	Speech	pov*	*There was some **pov**erty of **speech** and content of thought.*
Optional	hallucinat*	Audit*	*For past 1 week has been having **audit**ory command **hallucinations** telling him to kill himself and also suicidal ideation.*
None	Rapport		*When she was last seen at her CPA on XXXXX by Specialist Registrar Dr XXXXX, ZZZZZ presented as well kempt with good eye contact and **rapport**.*

A key task in CRIS development, which underlies a significant component of this thesis, has been to create data on a range of symptoms of mental disorders in order to 'unlock' this information and render it available for research use – e.g. being able to characterise a patient over a given time period on the basis of the mental health symptoms recorded during that period. In terms of the symptoms prioritised for inclusion, firstly, the potential salience of symptoms for research applications was considered. The Positive and Negative Symptoms Scale (PANSS) (Kay et al. 1987) and Young Mania Rating Scale (Lukasiewicz et al. 2013) were also used as templates for guidance in this respect. Secondly, the language used in routine clinical records was taken into consideration in choosing symptoms, focusing particularly on those which were likely to be recorded in the most consistent and tractable language, based on the experience of the clinicians within the group. Thirdly, it was decided *a priori* to extract sufficient numbers of symptom types to generate scales for further evaluation within the following domains: i) positive symptoms; ii) negative symptoms; iii) disorganisation symptoms; iv) manic symptoms; v) catatonic symptoms; vi) depressive symptoms (Table 5-2). The first four of these followed the findings of Demjaha and colleagues (Demjaha et al. 2009) at an early stage, who had found that factor analysis gave rise to a five-factor solution of manic, reality distortion, negative, depressive and disorganisation symptom dimensions. The scores of identified dimensions were differentially associated with specific variables. The manic dimension had the highest number of significant associations (strong correlations were observed with shorter duration of untreated psychosis, acute mode of onset and compulsory admission). Adding dimensional scores to diagnostic categories significantly increased the amount of variability

explained in predicting these variables; the reverse was also true but to a lesser extent. The study concluded that categorical and dimensional representations of psychosis are complementary and suggested using both could be a promising strategy in conceptualising psychotic illnesses. Depressive symptoms were looked at later. Catatonic symptoms were further added to improve consistency with the study of Cuesta and colleagues (Cuesta & Peralta 2001), and as a symptom group of interest which is often not adequately captured in dimensional studies because of its relative rarity in recruited clinical samples. Depressive symptoms were ascertained at a later stage of development, driven by a particular research priority at that time to develop an algorithm to ascertain treatment resistance.

Table 5:2- Symptom groupings with “key terms” applied for CRIS natural language processing

Domain	Key terms (symptoms)
Positive	Agitation, aggression, arousal, hostility, delusions, hallucinations, paranoia, persecution
Negative	Blunted or flat affect, emotional withdrawal, social withdrawal, abstract thinking, poor rapport, apathy, poverty of speech, poverty of thought
Disorganisation	Circumstantial speech, reduced coherence, formal thought disorder, thought block, tangential speech, derailment, flight of ideas
Manic	Elevated mood, disturbed sleep, insomnia, pressured speech, irritability, elation, grandiosity
Catatonic	Mannerism, rigidity, posturing, perseverance, stupor, waxy flexibility, immobility, echolalia, mutism, catalepsy, echopraxia
Depressive	Anergia, anhedonia, loss of appetite, poor concentration, low energy, guilt, helplessness, hopelessness, poor motivation, psychomotor retardation, suicidal ideation, tearfulness, weight loss, worthlessness

5.4.2 Personal contributions to application development and annotation of SMI symptom concepts

The development of NLP applications ascertaining symptoms of psychotic and affective disorders was a substantial task carried out as a team rather than by individuals. Having secured my studentship around the commencement of this initiative, it was envisaged that I would assist with the development and lead on specific analyses which exploited the new data in relation to physical health and mental health service use outcomes. This thesis therefore describes the development work to which I contributed and the analyses which I led. As will be discussed in later chapters, my involvement in the development work assisted a great deal in interpreting findings derived from it.

My key contribution was participating with colleagues in drawing up a list of potential symptoms of interest and in providing manual annotations of text from clinical records as training data for the algorithms. I provided input to developments to produce annotation guidelines to ensure consistent, high quality gold standard and training data, and annotation guidelines were developed by the team based around internal, iterative discussions. Developments to applications created and data extracted relevant to this thesis were carried out by a team of individuals including myself. Within the team, Richard Jackson, an informatician and fellow PhD student, led on TextHunter software development. Other team members undertook leading the annotations for the symptom domains as follows: positive symptoms were developed by myself; negative symptoms by Professor Robert Stewart (Consultant Psychiatrist and my Primary Supervisor); disorganisation symptoms shared between myself, Professor Robert Stewart and Dr Rashmi Patel (Clinical Research

Fellow and Senior Trainee Psychiatrist); manic symptoms by Dr Rashmi Patel; and catatonic symptoms by Dr Anna Kolliakou (Post-doctoral Research Associate). Subsequent to the main period of work on this, a range of depressive symptom NLP algorithms were developed by Professor Robert Stewart, assisted by Jyoti Jyoti and Dr David Chandran (Informaticians), and were incorporated in later research papers.

To optimise the performance of the language models for the SMI cohort, the training corpus was enriched by selecting any text occurrence in CRIS (irrespective of the document type), relating to a patient who had received an SMI diagnosis, defined as schizophrenia (ICD-10 code F20.x), schizoaffective disorder (F25.x) or bipolar disorder (F31.x). This diagnosis information was drawn from structured fields in the source EHR, which are completed by clinicians during the normal course of care by means of selecting an appropriate ICD-10 code. These codes were supplemented by a separate NLP application which returns searchable text strings associated with diagnostic statements in text fields. In UK NHS mental health Trusts, recording of diagnosis is effectively mandatory, but recorded diagnoses themselves have no financial implications for Trusts (e.g. are not used for any billing purposes).

When I carried out annotations, a relevant instance was defined as a mention of a symptom observed in a patient, without a grammatical negation which is where a term contradicting some or all of the meaning of an affirmative (positive) sentence was found. Independent sets of gold standard data were also created for each symptom to assess the performance

of each model. These were derived in the same manner as the training data which meant that annotations were carried out in a similar way for both sets of data. For both training and gold standard data, a relevant instance of a symptom was labelled as 'positive', (such as 'the patient had poverty of speech') whereas irrelevant or negation instances (such as 'today I examined the patient for poverty of speech…' or 'the patient did not have poverty of speech') were combined into a 'negative' category to create a binary classification problem. Once the agreed number of annotations were completed, the informatics team members were made aware. They would then use the training data to carry out a 10-fold cross validation to estimate the optimal SVM model parameters using the features provided by TextHunter. An instance was considered correctly classified if the sentence containing the human label of 'positive' or 'negative' and symptom type matched the model generated label and symptom type. Sub-classifications based upon any optional modifiers that were present (e.g. modality of hallucination) were not evaluated in this work. Finally, the optimised models were validated against the gold standard data. Arbitrarily, the gold standard for each concept was to contain a minimum of 100 'positive' mentions, to derive precision and recall measures for the 'positive' class. The findings were relayed to me (for positive symptoms and the specific disorganisation symptoms I annotated) as well as a sample of output from the initial prototype. I then used these outputs to stimulate discussion about requirements and to provide a basis for multiple iterations of development until performance requirements were met with senior colleagues. For several symptoms, at least within the positive symptom domain this usually meant additional annotations.

5.5 Data extraction from CRIS Linkages

CRIS has been linked to other local or national databases, such as Lambeth DataNet (local Primary Care), Hospital Episode Statistics, the Department of Education National Pupil Database, the national Cancer Registry and national mortality data. In this thesis, two data linkages were used: i) linkage between CRIS and the Office for National Statistics (ONS) mortality data, which includes information available on individuals' death certificates, including recorded cause of death; and ii) Hospital Episode Statistics where admissions data are compiled centrally for all Trusts in England, both acute and mental health services.

5.5.1 Mortality data

CRIS contains a limited level of mortality information for deceased individuals such as date of death (Perera et al. 2016). As used in chapter 6 and 9, to obtain more detail, anonymised BRC IDs are linked to the ONS death register using SLaM's Clinical Data Linkage Service (CDLS). The CDLS send identifiers (CRIS ID, first name, last name, date of birth, gender, postcode and NHS number) to NHS Digital, who in turn request the mortality data from the ONS, and then return the mortality data to CDLS via a secure file transfer protocol. The CDLS is responsible for linking, extracting and storing the data in accordance with the SLaM ICT Security Policy on a server within the SLaM firewall (Perera et al. 2016).

5.5.2 Hospital Episode Statistics (HES)

HES data are compiled centrally for all secondary care Trusts in England, both acute and mental health services. These data include statistical abstracts of records of all inpatient episodes, as well as outpatient and emergency care, and within the time CRIS has been operational were held initially by the Health and Social Care Information Centre (HSCIC) and then by NHS Digital. CRIS identifiers were compiled by CDLS, and transferred to HSCIC using an NHS-approved secure file transfer protocol. HSCIC then added the CRIS identifier to all HES records that match CRIS records and extracted all other HES records for residents within the four catchment boroughs served by SLaM- Lambeth, Southwark, Lewisham and Croydon (in chapter 7- the comparison group). HSCIC destroyed patient identifiers leaving only the CRIS ID and HES extract ID. As with other linked data sets, the CRIS-HES data were transferred back to CDLS to be held and provided to researchers in a fully anonymised format – i.e. replacing the CRIS identifier with a project-specific anonym. This means that HES episodes will refer to any hospitalisation in England for individuals on CRIS (in chapters 7 and 9) and for anyone resident in the catchment area in chapter 7.

5.5.3 Personal contributions to data extraction

Following discussions with supervisors, I prepared data collection templates that would then stimulate discussion with the clinical informaticians in the CRIS team who prepare the SQL query for data extraction. While the data extraction process was completed in collaboration with the clinical informaticians, I provided detailed instructions on the selection parameters, specifying individual variables, coding and time frames for measuring exposures and outcomes. Furthermore, I worked closely with clinical informatics colleagues to ensure most recent updates to applications developed were used for data extraction. Following extraction, I conducted all the data analyses and (with input from my supervisors and co-authors) prepared manuscripts for peer-review.

5.5.4 Study variables extracted from data sources

The analyses described in Chapters 6,7,8 and 9, primarily include individuals who had received a SMI diagnosis (the cohort in chapter 8 was only individuals with schizophrenia), who were aged 15 years or over at first diagnosis, and who were active to SLaM clinical services during the observation period. Each chapter defines the specific observation periods for each investigation. 'Active to SLaM services' was defined as having an open team episode in PJS: in other words, an on-going contact between the patient and a team within the service. SMI was defined as any previous receipt of a schizophrenia (ICD-10 code: F20.x), schizoaffective disorder (F25.x) or bipolar disorder (F31.x) diagnosis. All three conditions represent clinically significant psychiatric syndromes, relatively unlikely to be treated without hospitalisation at some point in a lifetime. Thus, they have a high 'penetration' to secondary mental healthcare records. Diagnostic data were derived from diagnostic structured fields within CRIS supplemented by the NLP diagnostic algorithms. For individuals identified as having received a SMI diagnosis, their first diagnosis was used throughout the time frame. This means that a second diagnosis or multiple diagnoses at multiple visits was not considered which meant a consistent approach was followed but accuracy of information may be lowered. However, diagnostic categories are expected to overlap since a proportion of individuals will suffer from more than one mental disorder (Chang et al. 2010). More specifically, Chang et al 2010, utilising CRIS as its data source found that of the sample, 1,680 (5.3%) had two diagnoses of interest before the end of 2009, 121 (0.4%) had three diagnoses and six (0.02%) had four diagnoses. In this study, the most common combinations of diagnoses applied on different occasions in the same individual were schizophrenia and schizoaffective disorders (n = 421). Only individuals with

at least one diagnosis of schizophrenia, schizoaffective disorder or bipolar disorder were included, based on assigned ICD-10 classification. Other psychotic conditions, for example, F21-Schizotypal disorder, F23-Acute and transient psychotic disorders or F29-Unspecified nonorganic psychosis, have more variation in duration and severity (World Health Organization 2010), making any estimates less reliable as well as not reflective of SMI.

5.5.5 Main exposures and outcomes

Main exposures and outcomes for each analysis are described in relevant chapters. In this chapter, I have described the covariates that were included in each of the analyses.

Age

Age was derived from a truncated version of the patient's date of birth (truncated to year and month to avoid identification), which is a structured field in PJS and was calculated at a defined time point as specified in each analysis chapter. inclusion age ranges for the different chapters varied and will be described for each analysis.

Gender

Gender was derived from a structured field in CRIS and classified as either male or female. Gender information was available for all individuals included in the analyses.

Ethnicity

Ethnicity was extracted from CRIS. In SLaM records, ethnicity is classified according to standard census codes and was available as one of the following and pragmatically condensed into groups- White (British, Irish, any White background),Black (African, Caribbean, any Black background), Asian (Indian, Chinese, any Asian background)and Mixed/ unknown/ missing.

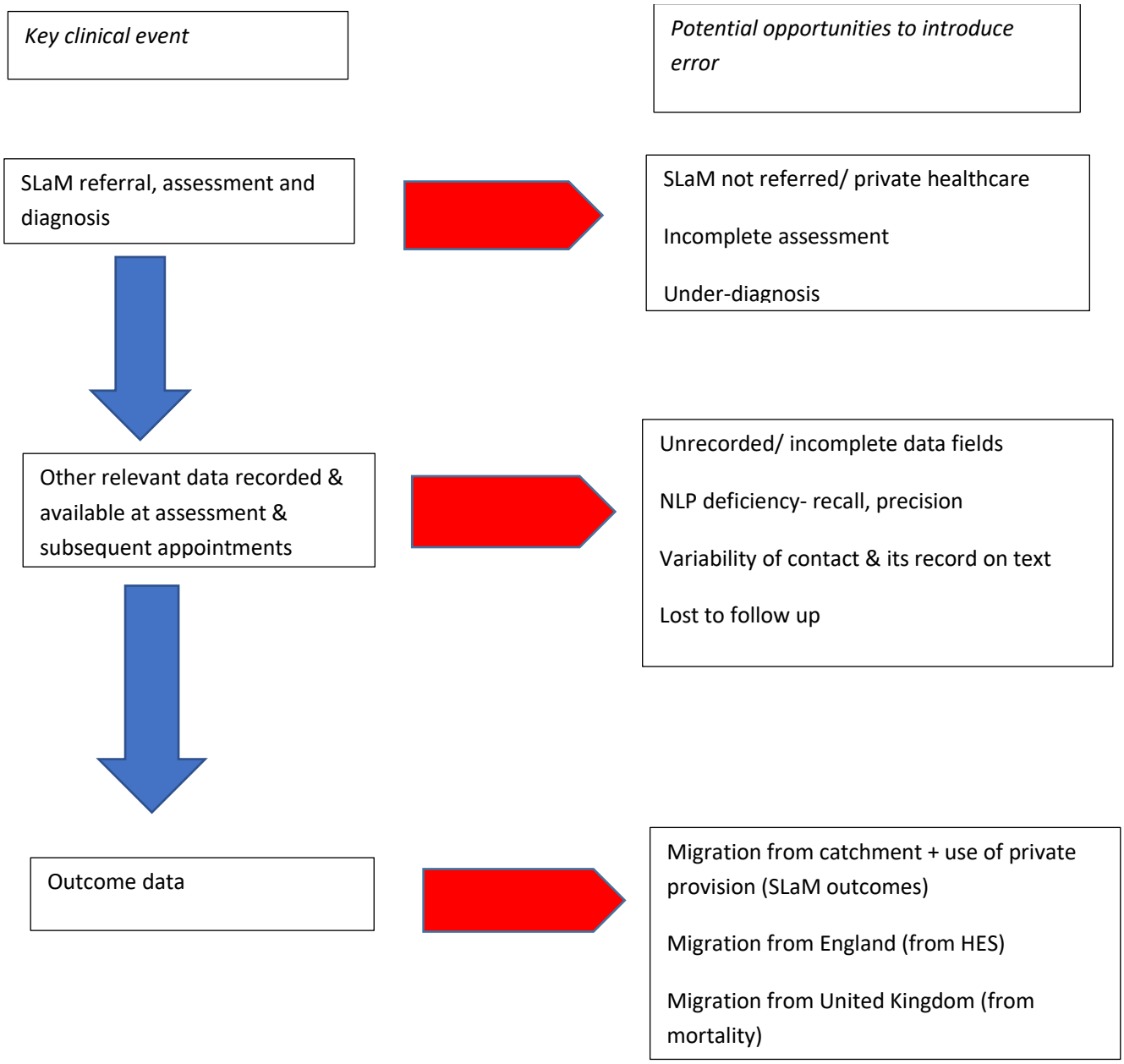
Relationship status

Relationship status was derived from structured fields and has a total of eight categories- cohabitating, married, in civil partnership, single, divorced, separated, widowed, unknown.

Employment status

Employment status was derived from a structured field. It contained sixteen categories, which were employed; other employment status; paid employment; part-time employment; self-employed; full-time student; full-time student school age; not known; not applicable; not disclosed; not known; other; registered disabled; retired; unemployed; volunteer.

Figure 5-1- Data Journey and potential opportunities for introduction of errors



Data linkages to mortality and hospitalisation

Using nationally available data sources such as Office for National Statistics mortality data as well as Hospital Episode Statistics meant that mortality data and hospital admission data are as comprehensive as possible. Completeness is good for both these data sources as there is a legal requirement to record every death while the hospitalisation data have an administrative function and are linked to payments to hospitals. The linkage to a national data source also meant that loss to follow up is very limited unlike that described in other studies (Fiedorowicz et al. 2009). There are limitations associated with the linked data and these include the quality of cause of death information on death certificates, and the quality of discharge diagnosis data on HES. There are potentially some limitations on coverage. For example, mortality data on Office for National Statistics are only ascertained in the UK and HES data are only collected in England covering a large geographical distribution (figure 10.1). Therefore, this needs to be borne in mind when interpreting findings on different ethnic/nationality groups. For example, my data source would not have picked up a death in an individual who had returned permanently to the Caribbean or South Asia.

Dealing with deficiencies in diagnostic distinctions

In chapter 1, I referred to literature citing the need for a more dimensional approach because SMI as a cluster may have more validity than its specific component diagnoses. As discussed, this was a key justification for developing measures of symptom dimensions. This is also pertinent to considering clinician variability (figure 10.1) in diagnoses entered onto the EHR, codes on death certificates, and codes on discharge diagnoses, combined with the

need to conform to ICD-10 classification. It is likely that there will be more agreement on broader than narrower diagnostic groupings, hence the approach taken in my thesis. However, embedding more broader groupings that are relatively heterogeneous categories means the more nuanced effects defining the specific component diagnoses will be overlooked.

Use of NLP to improve characterisation

The problem of sub-diagnosing SMI supports the aim to improve characterisation through NLP. This approach undoubtedly has helped to extract data from unstructured fields on large samples which would otherwise not have been feasible. Information extraction processes focusing on free text which would include correspondence maximises the generalisability of the approach, as letters to primary care physicians, which accounted for a large portion of the correspondence text, are unlikely to vary substantially between mental health services with respect to the language used to describe the symptoms of interest.

The key metrics traditionally used in NLP for evaluating the information extraction application performance in the test data sets, at the level of the individual text annotation are precision and recall. These are, respectively, the proportion of application 'hits' which are found to identify the genuine construct and the proportion of instances of the genuine construct which are identified by the application. Preference to maximise precision over recall impacts the latter thereby introducing error. Error can also vary by symptom as some terms occur more commonly in other contexts too such as "motivation".

It is possible to search the patient's entire record but it was not possible to delineate the timing or duration of individual symptoms, or whether they were primary (for example, a direct consequence of illness) or secondary (for example, side effects of treatment) as this is not routinely documented. However, for the former at least, by setting end points to observation periods and start dates for follow up, it has been possible to measure temporal relationships. In this thesis, the extent to which individual symptoms could be considered as having equal weight in a composite score was not explored. The application of NLP to clinical records is clearly unlikely to identify symptoms as accurately as a direct assessment using a specialised psychopathological rating scale, and consistency of assessment and recording is likely to be lower in clinical staff than trained interviewers. Underestimation of symptoms may have occurred for several reasons. First, because a minimum length of treatment was not specified in the inclusion criteria, relatively new patients with sparse documentation may not have much symptoms information in their record. Second, aims to improve precision over recall in the models may have reduced the probability of detection. Third, the list of symptoms used was not comprehensive and may have missed some aspects of psychosis presentation – either because of different symptoms which were missed, or because of target symptoms which were described in non-standard language (for example, 'hearing voices', rather than 'auditory hallucination'). A final consideration regarding symptoms data is that the appearance of a given term in the record depends on the perceived salience of that term in routine clinical practice. A symptom which is known to reflect an individual's state of health (e.g. sleep disturbance, grandiosity, paranoia) may be frequently monitored and thus recorded, whereas others may be less noticeable and less

likely to be recorded. This may in turn depend on other factors. For example, even though schizophrenia and bipolar disorder are recognised to be overlapping entities, the presence of a particular diagnosis may well influence the symptoms a clinician particularly looks for and chooses to record; likewise, there may be influences arising from a patient's social class or cultural background which have a similar effect.

5.6 Statistical analysis

In chapter 6, I describe in more detail the use of life tables to estimate life expectancy. For estimations, I used Chiang's abridged life-table method (CL 1984) on Microsoft Excel software and data available from the Office for National statistics which included the England & Wales population estimates. Death rates in the SMI cohort for each major primary cause category were sequentially replaced with England and Wales gender- and age-specific rates for that cause of death, obtained from UK Office of National Statistics.

In chapter 7, using STATA 12 and Microsoft Excel software, age- and sex-standardised admission ratios (SARs) were calculated for the two years of the observation period for the cohort with SMI. SARs were calculated first for diagnoses grouped at the highest level (letter) ICD-10 codes relating predominantly to the body system affected or type of disorder Analyses were also carried for more specific three-character diagnostic codes, describing those accounting for at least 1% of all hospitalisations in the SMI cohort.

In chapter 8, I used STATA 11 software, to describe the distribution of negative symptoms, and to investigate the associations between the clinical outcomes and the presence of negative symptoms, using binary logistic regression; and symptom scale scores, using ordinal logistic regression. Associations between negative symptomatology and hospital admission and readmission were analysed using logistic regression, while those with length of inpatient stay were analysed using linear regression. For the analyses with hospitalisation outcomes in/following 2011. Secondary analyses were undertaken to investigate and

compare the relationships of individual negative symptoms of the CRIS- Negative Symptom Scale with risk of readmission and length of stay using binary logistic and linear regression, respectively.

In chapter 9, using STATA 13 for the analyses on of mortality as an outcome within the cohort, univariate and multivariate analyses were carried out for the effect of specific symptom domains of SMI on all-cause mortality, using Cox regression with the adjustment for age, gender, ethnicity, employment status and marital status as potential confounders. For the analyses on of hospitalisation outcomes, univariate and multivariate linear regressions were applied individually for hospital admissions by symptom domains.

Details for the specific analyses are available here: Chapter 6 section 6.3.3 and 6.3.4; Chapter 7 section 7.3.4; Chapter 8 section 8.3.4 and Chapter 9 section 9.2.3.

The following four chapters which describe the results are based on published/ in submission papers. I have altered the wording to indicate my personal role in the papers and have added explanatory summaries. Within Methods sections where some information repeated itself from chapter to chapter, example- data sources, I have excluded the paragraphs in the latter chapters and refer to former chapters or chapter 5.

6 Contributions of specific causes of death to lost life expectancy in severe mental illness

6.1 Chapter outline

This chapter investigates the major disease groups that contribute to the gap in life expectancy between individuals with severe mental illness and the general population. Given the life expectancy gap between people with severe mental illness and the general population persists and may even be widening (Lawrence D, et al 2013), this chapter aims to estimate contributions of specific causes of death to the gap.

The chapter has been published in the following peer-reviewed paper-

Jayatilleke N, Hayes RD, Dutta R, Shetty H, Hotopf M, Chang CK, Stewart R. Contributions of specific causes of death to lost life expectancy in severe mental illness. European Psychiatry. 2017 Mar 17;43:109-115.

Personal contributions to publication- Using concepts discussed with supervisors, I designed the study approach. I had full access to all the data in the study and took responsibility for data integrity, accuracy of the data analysis, interpretation of data, drafting and critical revisions of the manuscript.

6.2 Research question

The life expectancy gap between people with severe mental illness (SMI) and the general population persists and may even be widening. This analysis aimed to estimate contributions of specific causes of death to the gap.

6.3 Background

Premature death among people with mental disorders is an international concern (Chesney, E, Goodwin, GM, Fazel 2014). Individuals with severe mental illness (SMI), including schizophrenia, schizoaffective disorder and bipolar disorder, experience particularly higher all-cause mortality and lower life expectancy (Laursen TM, Munk-Olsen T, Nordentoft M 2007; Chang CK, Hayes RD, Perera G, Broadbent MT, Fernandes AC, Lee WE, Hotopf M 2011). This is an important indicator for policy development to improve on current health inequality; however, contributions of individual causes of death to the life expectancy gap remain unclear, limiting the evidence. Previous research has suggested excesses of both natural and unnatural causes of death, and that the excess of natural causes may cross several mental disorder groups (Hoang et al. 2011). Nevertheless, standardised mortality ratios cannot be translated into the more important construct of life expectancy which accounts for the age of death as well as the absolute risk. Specifically, life expectancy as a measure adds value because it is a demographic index which is a feature of overall mortality emphasizing the impact of deaths occurring in younger age groups. It can thus be claimed to have considerably more 'meaning' for a population as an indicator of disadvantage, compared to a relative risk or hazard of mortality over a defined period of time. Life expectancy as a measure is obtained from estimates from the age-specific mortality of a specific cohort over a given period of time using the life table method and is calculated from the accumulated person-years contributed by the entire cohort divided by the total population number at birth.

The substantial disadvantage in life expectancy experienced by people with SMI represents a major health inequality. Recent studies have shown how the impact of chronic illness varies between people with SMI and the general population. For example, the relative risk of Type 2 diabetes is elevated in younger populations of SMI compared to general population. In one study, most associations persisted despite adjustment for anti-psychotic prescriptions, and ethnic minority groups also had a higher prevalence of Type 2 diabetes in the presence of severe mental illness (Das-Munshi et al. 2016). On a similar note, the prevalence of multi-morbidity (which often encompasses combinations of mental and physical illnesses) is recognised to increase with higher deprivation (Barnett et al. 2012). Further, in people with SMI, prevalence of obesity, metabolic syndrome, diabetes mellitus, symptoms of cardiovascular disease, and respiratory disease all exceed those in the general population at least two-fold, and HIV prevalence may be increased as much as eight-fold (Scott & Happell 2011). This increased prevalence of chronic disease may be largely responsible for a five-fold increased risk of death (Scott & Happell 2011).

An analysis was strongly warranted to add clarity to this 'survival gap', particularly as interventions have been proposed which focus on individual pathways without clear evidence on the contribution of those pathways to the overall outcome. This chapter is a descriptive analysis aiming to explain the above. Using data from a large clinical cohort, I estimated the contribution of the major groups of causes of death to shortened life expectancy in people with SMI, delineating to what extent this gap could conceivably be closed if mortality rates for each cause of death could be equalised with those in the general population.

6.4 Methods

6.4.1 Data sources and cohort

The South London and Maudsley NHS Foundation Trust (SLaM) is one of Europe's largest secondary mental healthcare providers, serving a single geographic catchment of approximately 1.36 million residents in four southeast London boroughs. In 2006, fully electronic health records were implemented across all SLaM services, and, in 2008, the Clinical Record Interactive Search (CRIS) system, supported by the National Institute of Health Research (NIHR) Biomedical Research Centre for Mental Health at SLaM, was developed to allow search and retrieval of anonymised but complete medical records within a robust patient-led governance framework (Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, Fernandes A, Hayes RD, Henderson M, Jackson R, Jewell A, Kadra G, Little R, Pritchard M, Shetty H, Tulloch A 2016; Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang CK, Jackson RG, Roberts A, Tsang J, Soncul M, Liebscher J, Stewart R 2013). Using CRIS, individuals who had a primary diagnosis of SMI (schizophrenia, ICD-10 coded F20; schizoaffective disorder, F25, or bipolar affective disorder, F31) and had at least one active contact with SLaM during the period between 1st Jan 2007 and 31st Dec 2012 were included in this analysis. Recorded primary diagnoses in structured fields of the record were supplemented by a natural language processing application developed using Generalised Architecture for Text Engineering (GATE) software to identify diagnostic statements in open text fields (e.g. correspondence and clinical notes) (Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang CK, Jackson RG, Roberts A, Tsang J, Soncul M, Liebscher J, Stewart R 2013; Taylor et al. 2015).

6.4.2 Mortality

Date and primary cause of death for study subjects in the same period (i.e. 2007-2012 inclusive) were retrieved from a CRIS linkage to death certifications, held by the UK Office of National Statistics which is updated regularly (Lopez-Morinigo JD,et al 2014), by NHS numbers through an anonymised process. The basic demographic information- gender (binary variable), date of birth and date of death were used to calculate age at death stratified by 5 year age bands (categorical variable). Cause of death was also a categorical variable where deaths due to Circulatory (I00-I99), Cancer (C00-D48), Respiratory (J00-J99), Digestive (K00-K93), Suicide (X60-X84, Y10-Y34), Other external (V01-Y98) and All other causes were distinct categories.

The same data source (i.e. UK Office of National Statistics), but this time via online data files, was used to obtain age-, gender, and cause-specific mortality for England & Wales population in 2010 (Office for National Statistics 2012).

6.4.3 Analysis: baseline life expectancy at birth

Using Microsoft Excel software, life expectancy at birth was estimated using Chiang's abridged life-table method (Chiang CL 1984). In brief, the accumulated person-years were calculated by gender based on age- and gender-specific mortality experienced by a specified male / female cohort in five-year age bands, and then divided by the total number of male / female cohort subjects at baseline. Because of the dynamic nature of the cohort during the six-year follow-up period (2007-2012), the number of deaths observed was weighted by the average at-risk period of each age and gender band. The life expectancy at birth for the England & Wales population using 2010 data was also estimated for the purpose of validation using Chiang's abridged method (Chiang CL 1984; Office for National Statistics 2012). Since SMI diagnoses are relatively unlikely to be given to people under the age of 15 years, mortality rates up to that age were assumed to be equal to those in the general population of England & Wales in 2010. The 95% confidence interval were found by multiplying the standard error by 1.96, obtained from the variance of life expectancy.

6.4.4 Analysis: re-calculated life expectancies by equalising mortality rates

The aim was to estimate the improvements in life expectancy which might be achieved if people with SMI experienced the same age- and gender-specific mortality rates for particular causes of death as those in the general population. For this a life expectancy simulation was used to examine the contribution of major causes of mortality to the reduced life expectancy in people with SMI. Death rates in the SMI cohort for each major primary cause category were sequentially replaced with England and Wales gender- and age-specific rates for that cause of death, obtained from UK Office of National Statistics (Office for National Statistics 2012). The expected numbers of deaths in each five-year age band were then re-calculated using these national gender- and age-specific rates. These re-estimations were carried out for cause-of-death categories by the order of number of deaths observed in the SMI study cohort: i) circulatory diseases, ii) cancer, iii) respiratory diseases, iv) digestive diseases, v) suicide, vi) deaths due to other (non-suicide) external causes, and vii) all remaining causes of death. In this study, in accordance with the ICD-10 manual, external causes comprised mortality involving injury, poisoning and certain other external causes. Causes of death in this broad group of scenarios include examples such as "accidental poisoning by amphetamine" (X41) or "driver of car injured in non-collision transport accident" (V48.5). Here, the underlying cause is classified according to the circumstances of the fatal injury, rather than the nature of the injury which is coded separately. Of note, this means that direct alcohol related mortality is considered under external causes. Where there was a reverse association identified for any cause-of-death category (i.e. a lower life expectancy for SMI subjects following equalisation of a specific cause-of-death category to general population), its contribution to loss of life expectancy

was set at zero. The percentage of person-year contribution to the life expectancy gap for each category of death was then calculated accordingly. In a sensitivity analysis, the adjustment for cancer was re-applied taking into account reference to cancer anywhere on the death certificate (i.e. rather than just as the primary cause of death). Finally, I carried out an additional stratification by affective (bipolar/schizoaffective) and non-affective (schizophrenia) SMI diagnosis in order to compare cause of death contributions between these subgroups.

6.5 Results

Calculated by Chiang's method, the life expectancy at birth for the England and Wales population in 2010 was 78.5 (95% CI 78.1-78.8) years for males and 82.4 (95% CI 82.1-82.7) years for females: identical to values previously reported by the UK ONS (Office for National Statistics n.d.). In the SLaM cohort, there were 1,558 deaths during the observation period among 19,106 subjects with SMI (816 deaths among 10,414 males and 742 deaths among 8,692 females)- table 6-1 shows the cohort characteristics. The life expectancy at birth in people with SMI was estimated to be 67.9 years (95% CI 67.1-68.6) for men and 72.2 years (95% CI 71.5-72.8) for women – i.e. reduced by 10.6 years for men and 10.2 years for women. The template abridged life table that was used for life expectancy calculations (using Chiang 1984 method) is shown in appendix 2.

Table 6:1: Demographic characteristics of study population and general population (n= 18,974 with 1,488 deaths)

Variable		Number of individuals (number of deaths)
Age		
	15-19	725
	20- 24	1,308
	25- 29	2,011
	30- 34	2,160
	35- 39	2,312
	40- 44	2,465
	45- 49	2,207
	50- 54	1,556
	55- 59	1,134
	60- 64	908
	65- 69	713
	70- 74	542
	75- 79	438
	80- 84	270
	85- 89	150
	90 above	75
Gender		
	Female-	8,630 (717)
	Male-	10,338 (771)
Employment status		
	Unemployed-	5,630 (410)
	In education or paid	2,003 (341)
	work-	8,157 (657)
	not known-	3,184 (80)
	Missing-	
Marital status		
	Single-	11,887 (695)
	Married/ in	2,538 (214)
	relationship-	1,931 (199)
	Separated/ divorced-	573 (191)
	Widowed-	2,045 (189)
	Not known-	
Ethnicity		
	White (British, Irish, any	9,754 (994)
	White background)-	
	Black (African,	5,730 (305)
	Caribbean, any Black	
	background)-	1,063 (67)
	Asian (Indian, Chinese,	2,427 (122)
	any Asian background)-	
	Mixed/ unknown/	
	missing-	

In the SLaM cohort, the median (interquartile range) ages of death by cause among females were as follows: cardiovascular diseases 73.6 years (62.5-82.6); cancer 67.7 years (55.1-77.6), respiratory conditions 77.3 years (68.1-85.3), digestive conditions 57.7 years (50.4-76.1), deaths due to suicide 45.8 years (28.5-56.6), deaths due to unnatural causes 52.3 years (37.4-64.9), and all other causes 75.4 years (51.6-84.8). Among males, median ages of death by cause were: cardiovascular diseases 65.4 years (55.5-74.7), cancer 69.9 (60.9-76.6), respiratory conditions 70.6 years (61.3-78.7), digestive conditions 54.8 (43.8-65.9), deaths due to suicide 39.5 years (30.6-48.8), deaths due to unnatural causes 43.3 years (32.1-54.5) and all other causes 56.5 years (41.7-75.0). Summarised in Table 6 1, the leading cause of death in people with SMI was circulatory diseases, followed by cancer, respiratory disease, gastrointestinal disorders, suicide, and other external causes.

Figures 6.1 and 6.2 display the improvements to life expectancy by gender when major cause of death categories were replaced with England and Wales 2010 population rates. The incremental changes to hypothetical life expectancy of men with SMI from the baseline 67.9 years were: 69.7 years by equalising circulatory disease death rates to those in the general population, 69.7 years (i.e. no change) by further equalising cancer mortality, 71.4 years equalising respiratory disease, 72.3 years equalising gastrointestinal diseases, 73.6 years equalising suicide, 74.5 years equalising all other external causes, and finally attaining the national level of 78.5 years by equalising all remaining causes. The corresponding changes to the life expectancy for women with SMI from the baseline 72.2 years were: 74.4 years (circulatory diseases), 75.2 years (cancer), 76.6 years (respiratory diseases), 77.7 years (gastrointestinal diseases), 79.1 years (suicide), 79.8 years (other external causes), and 82.4 years (all remaining causes). These increments are expressed in terms of life years gained

and as proportions of the life expectancy gap in Table 6.1. Among ‘all remaining causes of death’, of note, infectious diseases (A00-B99) contributed 1.12 years to the life expectancy gap in males and 1.73 years in females.

Figure 6-1- Estimated contributions of different causes of death to the loss of life expectancy of men with SMI

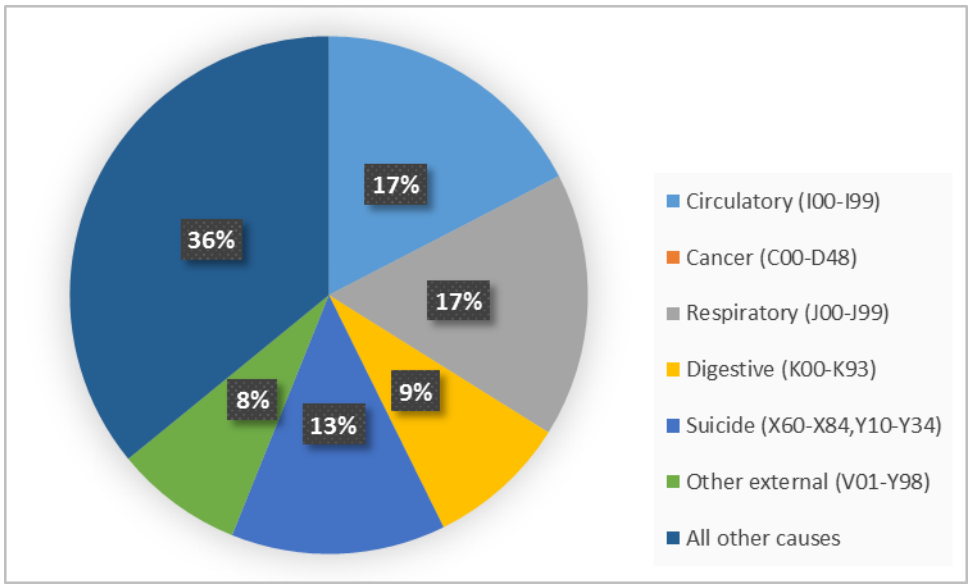


Figure 6-2- Estimated contributions of different causes of death to the loss of life expectancy of women with SMI

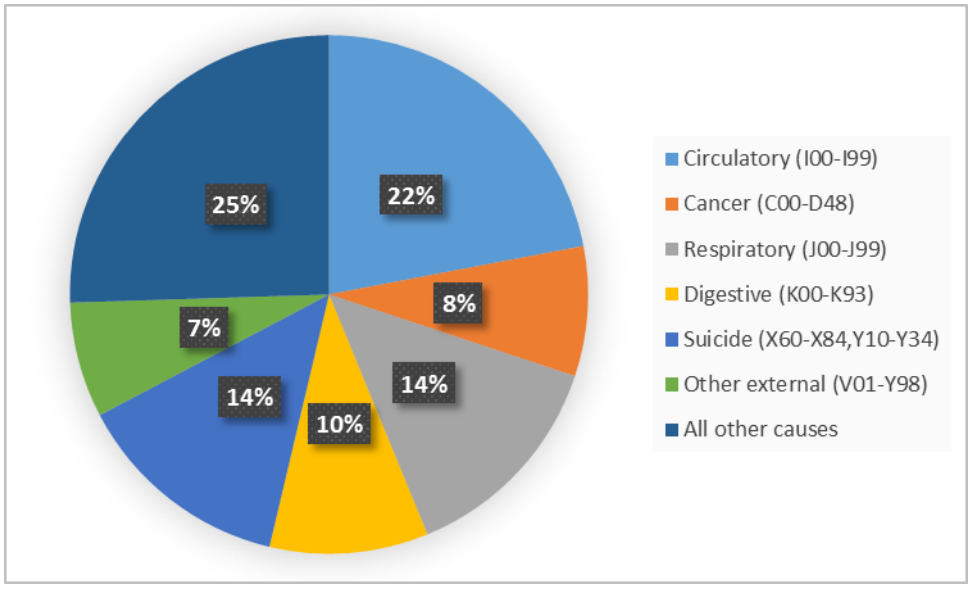


Table 6:2- Number of deaths by primary cause in people with SMI and the hypothetical change in life expectancy following equalisation of these to England and Wales cause-specific mortality rates in 2010

Gender	Primary cause of death (ICD-10 code)	Number of deaths during 2007-2012 (% of all deaths)	Life expectancy improvement in years by equalising causes of death with those in general population	Percentage contribution to the overall life-expectancy gap
Males	Circulatory (I00-I99)	195 (23.9)	1.82	17.4
	Cancer (C00-D48)	110 (13.5)	-0.02*	--
	Respiratory (J00-J99)	132 (16.2)	1.73	16.5
	Digestive (K00-K93)	55 (6.7)	0.92	8.8
	Suicide (X60-X84,Y10-Y34)	55 (6.7)	1.40	13.4
	Other external (V01-Y98)	57 (7.0)	0.84	8.0
	All other causes	212 (26.0)	3.76	35.9
Females	Circulatory (I00-I99)	205 (27.6)	2.24	22.0
	Cancer (C00-D48)	148 (19.9)	0.83	8.1
	Respiratory (J00-J99)	114 (15.4)	1.40	13.7
	Digestive (K00-K93)	40 (5.4)	1.01	9.9

Gender	Primary cause of death (ICD-10 code)	Number of deaths during 2007-2012 (% of all deaths)	Life expectancy improvement in years by equalising causes of death with those in general population	Percentage contribution to the overall life-expectancy gap
	Suicide (X60-X84,Y10-Y34)	33 (4.4)	1.39	13.6
	Other external (V01-Y98)	28 (3.8)	0.74	7.2
	All other causes	174 (23.4)	2.60	25.5

* Replaced by zero in consecutive calculations

Of all cancer mortality in women with SMI, the commonest cause in the sample was breast cancer (n=179) with a mean age at death of 67.7 years, followed by lung cancer (n=110) with a mean age at death of 71.7 years. For males, the commonest cancer-causing death was lung cancer (n=128) with mean age at death of 70.9 years followed by prostate cancer (n=101) with mean age at death of 78.0 years. A sensitivity analysis was carried out to address the issue of cancer as an underlying but not primary cause of death, equalising mortality rates where cancer was recorded anywhere on the death certificate. In summary the findings were not changed substantially, with the adjustment resulting in a 0.06 year reduction of life expectancy among males with SMI and an improvement of life expectancy for females by 0.22.

I then stratified the cohort by affective and non-affective subgroups, and Table 6.2 compares the contributions of the different causes of death to the life expectancy gap. For women, cancer and other causes were noted to contribute more in those with non-affective than affective disorders, while suicide, respiratory and digestive disorders contributed more in those with affective disorders. In men, respiratory disorders contributed more in non-affective disorders and other causes more in affective disorders. Other proportional contributions were similar between affective and non-affective groups.

Table 6:3- Contributions of individual causes of death to loss of life expectancy in SMI: stratification by affective vs. non-affective diagnoses.

Primary cause of death	Life expectancy improvement in years by equalising individual causes of death with those in general population (percentage contribution to the overall gap in life expectancy)			
	Males		Females	
	Affective disorders (202 deaths)	Non-affective disorders (569 deaths)	Affective disorders (299 deaths)	Non-affective disorders (418 deaths)
Circulatory (I00-I99)	1.00 (13.5)	1.03 (15.0)	2.25 (25.0)	2.29 (26.3)
Cancer (C00-D48)	0 (0)	0 (0)	0.23 (2.6)	1.22 (14.0)
Respiratory (J00-J99)	1.16 (15.7)	1.50 (21.9)	1.49 (16.5)	1.02 (11.7)
Digestive (K00-K93)	0.66 (8.9)	0.67 (9.8)	1.08 (12.0)	0.53 (6.1)
Suicide (X60-X84,Y10-Y34)	1.13 (15.3)	0.95 (13.9)	1.70 (18.9)	0.70 (8.0)
Other external (V01-Y98)	0.40 (5.4)	0.50 (7.3)	0.50 (5.5)	0.75 (8.6)
All other causes	3.05 (41.2)	2.20 (32.1)	1.76 (19.5)	2.21 (25.3)

6.6 Discussion

In a large cohort of people with SMI who had received routine mental healthcare, the contributions were investigated of different grouped causes of death to the reduced life expectancy, which itself is already well-recognised (Chang CK, Hayes RD, Perera G, Broadbent MT, Fernandes AC, Lee WE, Hotopf M 2011). The aim was to estimate the improvements in life expectancy which might be achieved if people with SMI could experience the same age- and gender-specific mortality rates for particular causes of death as those in the general population. The main finding was that the gap in life expectancy was accounted for by a wide range of recorded causes of death, rather than just one or two predominating ones. In addition, there was moderate variation in the pattern of contribution between men and women, and between people with affective and non-affective SMI diagnoses.

When assessing the specific contribution from each cause of death to life expectancy, it was difficult to identify a single group or two. As single groups of causes, circulatory and respiratory were highest in males and females, followed by cancer, digestive disorders, external causes and suicide; however, a tail of less common individual causes of death still contributed to 35.9% and 25.5% of the life expectancy gap for men and women respectively: greater contributions than cardiovascular disease and illustrating the high levels of heterogeneity in factors underlying mortality inequalities.

Considering the national context, in the general population during 2012, the last year of this analysis period, there were 499,331 deaths registered in England and Wales (Office for

National Statistics 2013). Age-standardised mortality rates in 2012 were 6,191 deaths per million population for males and 4,510 deaths per million population for females. In 2012, cancer was the most common cause of death (29% of all deaths registered) followed by circulatory diseases, such as heart disease and strokes (28% of all deaths registered). Of note, since 2002 death rates from cancer in the general population have fallen by 14% for males and 10% for females, and between 2002 and 2012, the male and female age-standardised death rates for circulatory diseases fell by 44% for males and by 43% for females (Office for National Statistics 2013).

As stated, life expectancy is a more meaningful construct than standardised mortality for public health, as it takes into account not only the level of mortality but also the age of death. Disorders affecting mortality at younger age therefore have a higher impact on life expectancy, even if overall associated mortality rates are similar or even lower than those of later-life disorders. Estimated life expectancies for male and female cohort members in this analysis were comparable to those previously reported from the CRIS database (Chang CK, Hayes RD, Perera G, Broadbent MT, Fernandes AC, Lee WE, Hotopf M 2011). Both are substantially lower than the life expectancy at birth for men and women in the England & Wales population in 2010. The importance of 'natural' causes (i.e. rather than suicide or violence) accounting for the mortality gap is also widely recognised. For example, from databases covering the state of Western Australia, in a population of 292,585 people in contact with mental health services, 77.7% of excess deaths were found to be accounted for by physical conditions, of which cardiovascular illnesses contributed 29.9% and cancers contributed 13.5%, compared to 13.9% of excess deaths caused by suicide (Lawrence D, Hancock KJ 2013). In that study, excess mortality by cause was defined as the difference

between the observed number of deaths (the total person-years in the cohort of mental health service users) and the expected number of deaths in the cohort (by major cause of death using cause-specific death rates by age group, sex, and time period) (Lawrence D, Hancock KJ 2013). Kessing *et al* (2015) also found that 74% of life years lost after age 15 was due to natural causes among men with SMI, and 80% in women (Kessing LV, Vradi E, McIntyre RS 2015). In this cohort, natural causes accounted for 78% and 77% of deaths respectively. Suicide clearly remains an important adverse outcome in SMI with a 5% lifetime risk estimated for people with schizophrenia (Palmer BA, Pankratz VS 2005), and accounted for 1.4 life years lost for both genders in this sample: contributing 13.4% of life-years lost in men and 13.6% in women. Of note, this contribution in women was stronger for those with affective compared to non-affective SMI diagnoses, but this diagnostic difference was not evident in men.

The findings revealed a relatively minor contribution of deaths from cancers, especially for men with SMI and women with affective SMI diagnoses. Other studies have also shown increase in mortality due to cardiovascular diseases and infectious diseases in SMI; however, unlike in this study, also indicated up to 1.4 to 2.0 fold increases in deaths from cancer (Nordentoft M, et al 2013). One potential explanation is simply that cancer deaths happen later than other causes, resulting in less impact on life expectancy among people with SMI, particularly if competing causes of death supervene – e.g. if a person at high risk of cancer dies from cardiovascular disease at an earlier age before the cancer has time to develop. From literature searches, it was clear that this study remains unique in its methodology emphasising the impact of mortality patterns for each major cause of death category, rather than comparing summed risks as applied elsewhere. This study additionally provides an

overview of all cancer-associated deaths in people with SMI in the given time period, supplementing previous findings of similar patterns in presentation / diagnosis of cancer to those in the general population (Chang C-K, et al., 2014). Although 13% of deaths for this cohort were due to cancer, the third leading cause of death, there was no improvement to life expectancy at birth among men when England and Wales cancer mortality rates were substituted. I considered the possibility that cancer was a potential underlying cause of death but not always considered or recorded as the primary cause in death certificates (e.g. because of deaths from respiratory disease in people with advanced cancer). However, there was minimal alteration in findings when mortality rates were substituted where cancer was recorded anywhere on the death certificate. The younger age of deaths from breast cancer could therefore explain the greater impact on life expectancy among females, particularly those with non-affective disorders. The finding for cardiovascular disease accounting for a higher proportion of life expectancy in women was similar to those of another study findings with higher contributions from cardiovascular disease to excess deaths in female compared to male patients with mental disorders (35.2% and 26.2%, respectively) (Lawrence D, Hancock KJ 2013).

Despite the well-recognised importance of deaths from natural causes in SMI, investigations and suggested interventions have tended to focus on specific causes of death such as cardiovascular disease, even though comparable morbidity has been reported in other systems (Partti K et al 2015). The findings confirm that this is likely to be over-simplistic and that a range of recorded causes potentially underlie the gap in premature mortality. The widening mortality gap between people with and without SMI is primarily accounted for by progressively improving survival rates in the general population which are not being

experienced by people with SMI, rather than by rising absolute mortality in SMI (Hoang et al. 2011). Furthermore, risk factors for mortality have high prevalence in people with SMI include smoking, alcohol, illicit drug use, poor diet and lack of exercise (De Leon J 2005; Lawrence D, Mitrou F 2009; Osborn DPJ, Nazareth I 2007; Kendrick T 1996; Davidson S, Judd F, Jolley D, Hocking B, Thompson S 2001). These are potentially implicated in a number of causes of death and their associations with mortality have been found to be stronger than expected in SMI (Davidson S, Judd F, Jolley D, Hocking B, Thompson S 2001). Excess mortality and morbidity associated with mental illness are recognised as multifaceted problems which are complicated to unpick (Lawrence D, Hancock KJ 2013). Higher levels of substance use, for example, have been well documented in people with mental disorders, including tobacco, alcohol, and illegal drugs, as well as a higher prevalence of unhealthy lifestyles, including poorer diets and less exercise (Lawrence D, Hancock KJ 2013). These findings suggest that higher incidence of illnesses associated with lifestyle risk factors might reduce life expectancy, supported by the fact that the majority of excess deaths in people with mental disorders are due to physical health conditions. It is likely that reducing mortality rates in the general population from common physical health conditions, such as heart disease, respiratory disease, and some cancers have not been experienced by individuals with severe mental illness to the same extent (Lawrence D, Hancock KJ 2013). General health promotion activities are therefore most likely to have an impact on life expectancy, including improvements in the availability and accessibility of prevention and risk management services, as well as addressing inequalities in treatment receipt where life-threatening conditions are present (Wu S-I et al 2013). It is also clearly important to consider side effects of some psychotropic medications, such as the weight gain and hyperlipidaemia which affect a high proportion of patients during initiation or maintenance

of antipsychotic treatment (Kendrick T 1996). Improving the availability and accessibility of prevention and risk management services could be an important means to minimise the mortality gap, but broader attention is also needed to address profound socio-economic inequalities which is more common in people with mental illness (Muntaner C, Eaton WW, Miech R 2004). Socio-economic inequalities are also associated with health risk behaviours (Muntaner C, Eaton WW, Miech R 2004); however, studies that adjusted for socioeconomic status still found significantly worse morbidity and mortality for people with mental illness, suggesting that social deprivation and disadvantage are not the sole determinants of poor health outcomes in these populations (Lawrence D, Hancock KJ 2013; Lawrence & Kisely 2010; Lawrence D, Jablensky AV, Holman CDJ 2000). Socio-economic inequalities are also associated with reduced access to healthcare and it is important to promote equality of access to good quality healthcare and evidence-based interventions for life-threatening conditions once these have developed, as a generalised deficiency in quality of healthcare might also underlie a heterogeneous group of conditions underlying loss of life expectancy.

Strengths of this study include the large sample size and linkage to nationwide mortality, with up to six years of follow up. This study adds to the literature by estimating the hypothetical contribution of each major cause of death to its current life expectancy compared to England and Wales which provides a fuller picture of the mortality gap than a comparison of absolute rates alone. Furthermore, this study also stratified the cohort into diagnostic subgroups within SMI. The generalisability of this analysis needs to be further evaluated, as the findings clearly refer to people with these mental disorders who had made contact with secondary mental health services within the specified time period. However, secondary healthcare setting is unlikely to bias findings to a large extent as SMIs by nature

are disorders where most cases will have received secondary care input. This is because in accordance with NICE clinical guidance individuals should be assessed by a psychiatrist or a specialist in psychosis and treated quickly if at risk of developing psychosis or have a first episode of psychosis (NICE 2014). The limitations of this study include its inclusion criteria, as its focus was on people who had received an SMI diagnosis (schizophrenia, schizoaffective disorder or bipolar disorder) whether or not they had other mental health diagnoses as well. I did not seek to investigate other conditions such as substance use disorders (F10-F19) and depressive disorders (F32- F33), although it is important to emphasise that they were not specifically excluded – i.e. comorbidity was allowable in the sample. The non-inclusion of pure (non-comorbid) substance use disorders may explain the slightly smaller gap between life expectancy observed in this study compared to a previous analysis (Chang CK, Hayes RD, Perera G, Broadbent MT, Fernandes AC, Lee WE, Hotopf M 2011); however, even without this group of individuals the overall gap in life expectancy remains extensive showing a very apparent need among those suffering from SMI. Although SLaM is a major academic as well as clinical centre, mental disorders defined from its records are likely to be broadly generalisable to other UK urban and suburban settings. An important consideration in drawing inferences from these findings concerns the assumptions underlying the analysis strategy, particular given it is a descriptive analysis and not hypothesis testing. These presupposed independent processes and did not take into account the potential for interactions between causes of death or competing risks of death (e.g. that equalising cancer mortality might have less impact than expected on life expectancy because of persisting cardiovascular diseases). The estimated contributions to the life expectancy gap therefore need to be viewed cautiously, although the principal finding of multiple contributions remains.

This study was carried out to improve understanding of premature mortality which has been a long-lasting problem among people with severe mental illness and described as "… at worst, a form of lethal discrimination" (Thornicroft G 2013). While we agree with the other opinion in this area that the time has come to shift efforts towards interventions to improve survival, we believe that observational data remain deficient as an evidence base and in need of improvement. In particular, the findings suggest that prioritised interventions should be those which demonstrably have an impact on multiple causes of death, since addressing a single cause will only have a small effect on overall life expectancy. In addition to receiving optimum mental healthcare, it is thus imperative that people with SMI should receive a package of care to help improve lifestyle, diet, and proper medication management. It is also important to bear in mind current thinking on achieving health improvements through reducing health inequalities and minimising the impact of socio-economic status on health advantages (Phelan JC, Link BG 2010). Failure to appreciate these challenges may result in resources being used inefficiently.

6.7 Summary of findings from chapter

Overall, natural causes (i.e. discounting suicide and other external causes) accounted for 79.2% (female) and 78.6% (male) lost life-years respectively. Deaths from circulatory disorders accounted for more life-years lost in women than men (22.0% versus 17.4%, respectively), as did deaths from cancer (8.1% versus no contribution), but the contribution from respiratory disorders was lower (13.7% versus 16.5%). More similar contributions were found for digestive disorders (9.9% and 8.8%), suicide (13.6% and 13.4%), and other external causes (7.2% and 8.0%). Loss of life expectancy in people with SMI is accounted for by a broad range of causes of death, varying modestly by gender.

7 Acute general hospital admissions in people with severe mental illness

7.1 Chapter outline

This chapter investigates the commonly recorded reasons for acute care hospitalisation in people with severe mental illness. Given that severe mental illness is associated with worse general health, reasons for admissions to general hospitals have received little investigation. This chapter aims to compare frequencies of non-psychiatric hospital admission by major diagnostic categories between people with SMI and the general population.

The paper reporting this chapter is currently under peer review

Jayatilleke N, Hayes RD, Chang CK, Stewart R. Acute general hospital admissions in people with severe mental illness. [In submission].

Personal contributions to publication- Using concepts discussed with supervisors, I designed the study approach. I had full access to all the data in the study and took responsibility for data integrity, accuracy of the data analysis, interpretation of data, drafting and critical revisions of the manuscript.

7.2 Research question

Severe mental illness is associated with worse general health, but reasons for admissions to general hospitals have received little investigation. This study aims to compare frequencies of non-psychiatric hospital admission by major diagnostic categories between people with SMI and the general population from the same geographic catchment.

7.3 Background

There is growing awareness of the adverse physical health experienced by people living with severe mental illness (SMI, which includes schizophrenia, schizoaffective disorder, and bipolar disorder). Specifically, people with SMI experience more medical comorbidities than the general population (Leucht et al. 2007; DE Hert et al. 2011), resulting in substantially higher mortality and shortened life expectancy (Chang CK, Hayes RD, Perera G, Broadbent MT, Fernandes AC, Lee WE, Hotopf M 2011; Lawrence D, Hancock KJ 2013). Nearly half of people with any mental disorder have comorbid medical problems such as diabetes, pulmonary disease or heart disease, and a further 35% have undiagnosed medical conditions, implying that only a small proportion are free of comorbidities (Zolnierek et al. 2009; Miller et al. 2006). As reviewed previously (Miller et al. 2006), this is particularly an issue for individuals with SMI, with the risk for diabetes being doubled for individuals with schizophrenia and trebled for individuals with bipolar disorder. Similarly, cardiovascular disease is increased 2-3 fold in prevalence, with particular impact on younger adults. Individuals with SMI are also at increased risk of pneumonia and need for mechanical ventilation, and higher risk of emphysema after controlling for smoking. In all, around 20% of people with SMI (Miller et al. 2006), bearing in mind there is high levels of under-reporting among individuals with SMI (Dorning, H. et al 2015), have been found to have a medical problem. Some of these conditions themselves may impact the mental condition by exacerbating signs and symptoms, response to psychotropic drugs, life expectancy and access to healthcare services.

People with SMI are thus known to be at high risk of a number of physical illnesses which cover multiple disease categories and body systems (Jones DR, Macias C, Barreira PJ, Fisher WH, Hargreaves WA 2004). However, there is no consensus on how to treat or prevent physical disease in this group (Mitchell AJ, Delaffon V 2012). In the UK, most people with SMI live in community accommodation and receive public universal healthcare in the same centres used by the general population (Health and Social Care Information Centre 2013). Levels of and reasons for the use of acute care (i.e. general hospital/secondary physical healthcare) services have not been adequately characterised for people with SMI, despite the importance of this information for primary care and other community preventative services, as well as acute care providers. Using a large linked database of mental health and acute care records, I sought to carry out a descriptive analysis of the commonly recorded reasons for acute care hospitalisation in people with SMI.

7.4 Methods

7.4.1 Sample and outcomes

The study setting has already been described in chapter 5-sections 5.2, 5.3 and in chapter 6-section 6.4. Cohort members were defined on the basis of a diagnosis of schizophrenia (ICD-10 code: F20.X), schizoaffective disorder (F25.X) or bipolar affective disorder (F31.X) recorded by SLaM on or before the 31st of December 2006 and on the basis of at least one contact with SLaM during 2007 or 2008. Diagnoses recorded in CRIS are based on the 10th edition of the World Health Organization International Classification of Diseases (ICD-10) and were extracted both from those recorded in structured drop-down menus within the source electronic health record or extracted from free text fields using bespoke software (Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, Fernandes A, Hayes RD, Henderson M, Jackson R, Jewell A, Kadra G, Little R, Pritchard M, Shetty H, Tulloch A 2016). Non-psychiatric hospitalisations were investigated over an observation period from 1st January 2009 to 31st December 2010 inclusive, using a data linkage between CRIS and Hospital Episode Statistics (HES) (Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, Fernandes A, Hayes RD, Henderson M, Jackson R, Jewell A, Kadra G, Little R, Pritchard M, Shetty H, Tulloch A 2016).

In the United Kingdom, HES data are compiled for all healthcare providers in England (both acute and mental health services), including statistical abstracts of records of all inpatient episodes, as well as outpatient and emergency care. A dataset was compiled and

standardised morbidity statistics calculated for people with mental disorders on the CRIS database in reference to the catchment general population living in Lambeth, Southwark, Lewisham and Croydon. Individuals who were under the age of 20 at start of the observation period or when they were firstly diagnosed as SMI during the period were excluded from analysis, and admissions to any mental health inpatient unit were excluded as outcomes.

Three-character ICD-10 codes for listed discharge diagnoses were extracted for each hospitalisation. A hospitalisation was defined as having at least one HES episode and contiguous episodes were combined where necessary into single hospitalisations. Age in the analysis was defined at the date of admission for physical illness where this occurred or at the mid-point of each one-year observation period for those not admitted in a specific year.

7.4.2 Diagnosis groupings and data management

The grouping of admission diagnoses was carried out in line roughly with the chapters of the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10), either by type of conditions (for example, cancers grouped together) or by organ affected, such as conditions of the eye (World Health Organization 2010).

For this study, data were obtained for the SMI cohort and comparator population using the CRIS-HES linkage. The age and gender profile for the SMI cohort was obtained on 1st January 2010 as the mid-point of the observation period for hospitalisations, and the age (in five year age bands) and gender profile of the comparator population was obtained from the 2011 UK census as this was the nearest to the observation period. Discharge diagnoses (one per hospitalisation; primary diagnosis if multiple) were obtained for the SMI cohort in 2009 and 2010. This study measures how much more likely or less likely a person with SMI is to be admitted to a general hospital compared to someone of the same age and gender living in the local area. Age and gender were accounted for through indirect standardisation and the generation of standardised admission ratios (SARs). An additional stratification was carried out by affective (bipolar/ schizoaffective disorders) and non-affective (schizophrenia) SMI diagnosis.

7.4.3 Statistical analysis

Age- and sex-standardised admission ratios (SARs) were calculated for the two years of the observation period (2009–2010) for the cohort with SMI using Stata 12. Using the observed number of specific admissions as the numerator, the denominator was the expected number of admissions in the same period of time, estimated using the age- and gender-specific admissions rates in 2009 and 2010 for SLaM's catchment area. SARs were calculated first for diagnoses grouped at the highest level (letter) ICD-10 codes: these relate predominantly to the body system affected or type of disorder, and were grouped/split as follows: i) splitting urinary (N00-N39) from other (N40+) genitourinary conditions; ii) combining neoplasms (all C codes and D00-D48) and separating these from blood disorders (D50+); iii) combining injuries and external causes (all S codes and T00-T14) and separating these from poisoning (T15+); iv) combining infections (all A and B codes). Analyses were finally carried for more specific three-character diagnostic codes, describing those accounting for at least 1% of all hospitalisations in the SMI cohort. I then excluded repeat admissions for the same ICD-10 three-character code for the disease groups and re-calculated ratios as sensitivity analyses.

7.5 Results

The SLaM SMI cohort consisted of 10,049 males and 8,331 females at the mid-point of the follow up period on 1st January 2010. The mean age of this cohort was 53.47 years. There were 8,622 admissions among individuals with a diagnosis of schizophrenia, schizoaffective disorder, or bipolar disorder who were observed from 1st January 2009 to 31st December 2010- 48.2% of these were admissions in males. In the comparison population (residents of the SLaM catchment areas) the total number of admissions was 501,158 among the 1.36 million population estimated from 2011 Census data (930,500 were aged 20 or over and included in the standard population). Table 7.1 displays numbers of hospitalisations by general chapters of ICD-10 codes in SMI cohort, crude admission rate in comparator population and age- and sex-standardised admission ratios in the cohort. The top five primary causes for hospitalisation were urinary conditions (ICD-10 codes: N00-N39), digestive conditions (K00-K93), unclassified symptoms (R00-R99), neoplasms (C00-D48) and respiratory conditions (J00- J99). SARs were raised for all of these apart from that for neoplasms where admissions were lower in the cohort. People with SMI also had higher risks of admission than the general population for poisoning and external causes, injury, mental and behavioural disorders, endocrine/metabolic conditions, blood disorders, nervous disorders, skin disorders and infections, as well as 'Z-codes' (another non-specific causes). Only neoplasms, musculoskeletal disorders, pregnancy-related disorders, and eye disorders showed significantly lower SARs. Figure showing-crude admission rates/ 1,000 persons in SLaM catchment area during 2009-2010 is shown in appendix 3.

Table 7:1- Age- and gender-standardised admission ratios (SARs) for hospitalisations in 2009-10 in people with severe mental illness (N of all admissions = 8,622), compared to the source population

Primary cause of admission (ICD-10 codes)	Number of hospitalisations in the SMI cohort	Crude admission rate in source population per 1,000 persons	SAR (95% CI)
Genitourinary system: urinary conditions (N00-N39)	1,886	84.4	1.60 (1.53- 1.67)
Digestive system (K00-K93)	974	58.9	1.51 (1.41- 1.60)
Symptoms, signs and findings, not elsewhere classified (R00-R99)	898	41.2	1.91 (1.79- 2.04)
Neoplasms (C00-D48)	632	69.7	0.66 (0.61- 0.71)
Respiratory system (J00-J99)	511	16.6	2.57 (2.35- 2.80)
Factors influencing health status and contact with health services (Z00-Z99)	453	31.0	1.72 (1.57- 1.89)
Circulatory system (I00-I99)	444	29.5	1.08 (0.98- 1.18)
Poisoning and other external causes (T15-T98)	368	6.2	5.03 (4.53- 5.57)
Injury (S00-T14)	352	14.6	2.52 (2.27- 2.80)
Mental and behavioural disorders (F00-F99)	299	0.9	30.3 (27.0- 34.0)
Musculoskeletal system (M00-M99)	280	27.7	0.77 (0.69- 0.87)

Primary cause of admission (ICD-10 codes)	Number of hospitalisations in the SMI cohort	Crude admission rate in source population per 1,000 persons	SAR (95% CI)
Endocrine and metabolic diseases (E00-E90)	277	7.5	3.07 (2.72- 3.45)
Pregnancy related (O00-O99)	256	72.4	0.84 (0.74- 0.95)
Blood disorders (D50-D89)	236	10.6	2.13 (1.87- 2.42)
Nervous system (G00-G99)	176	7.8	1.94 (1.66- 2.25)
Genitourinary system: pelvis, genitals and breasts (N40-N99)	165	18.6	0.98 (0.83- 1.14)
Eye conditions (H00-H59)	162	17.1	0.61 (0.52- 0.72)
Skin conditions (L00-L99)	159	10.1	1.54 (1.31- 1.80)
Infectious diseases (A00-B99)	78	3.8	2.06 (1.63- 2.58)
Ear conditions (H60-H95)	10	1.2	0.77 (0.37- 1.42)
Congenital abnormalities (Q00-Q99)	6	1.1	0.67 (0.25- 1.46)

Table 7.2 reveals re-calculated age- and sex-standardised admission ratios for the SMI cohort, excluding repeat admissions (defined as primary discharge diagnoses falling within the same grouped codes in the 2-year follow-up period). Here, the five most common reasons for admission, using general chapters of ICD-10 codes, were digestive, unclassified symptoms, respiratory conditions, 'Z-codes' and circulatory disorders. While the rankings were different from Table 7.1 in terms of the frequency of hospitalisations for SMI, the direction and significance of the SARs remained. Ratios of all admissions (Table 7.1) to admissions without including repeat admissions (Table 7.2) in people with SMI were highest for urinary conditions, neoplasms and blood disorders, indicating high numbers of repeat hospitalisations in those diagnostic groups.

Table 7:2- Age- and gender-standardised admission ratios (SARs) for people with severe mental illness excluding repeat hospitalisations (N of subjects with admissions = 5,377)

Primary cause of admission (ICD-10 codes)	Number of SMI subjects admitted (ratio of all admissions to unique admissions)	SARs (95% CI)
Digestive system (K00-K93)	780 (1.25)	1.37 (1.28- 1.47)
Symptoms, signs and findings, not elsewhere classified (R00-R99)	752 (1.19)	1.78 (1.65- 1.91)
Respiratory system (J00-J99)	409 (1.25)	2.55 (2.30- 2.80)
Factors influencing health status and contact with health services (Z00-Z99)	378 (1.20)	1.73 (1.55- 1.91)
Circulatory system (I00-I99)	347 (1.28)	1.06 (0.95- 1.18)
Injury (S00-T14)	331 (1.06)	2.52 (2.26- 2.81)
Mental and behavioural conditions (F00-F99)	299 (1.00)	30.3 (27.0- 34.0)
Poisoning and other external causes (T15-T98)	275 (1.34)	4.59 (4.07- 5.17)
Genitourinary system: urinary conditions (N00-N39)	257 (7.34)	1.67 (1.47- 1.89)
Musculoskeletal system (M00-M99)	237 (1.18)	0.80 (0.70- 0.91)
Pregnancy related (O00-O99)	217 (1.18)	0.80 (0.70- 0.92)

Primary cause of admission (ICD-10 codes)	Number of SMI subjects admitted (ratio of all admissions to unique admissions)	SARs (95% CI)
Endocrine and metabolic diseases (E00-E90)	205 (1.35)	3.49 (3.02- 4.00)
Neoplasms (C00-D48)	202 (3.13)	0.70 (0.60- 0.80)
Genitourinary system: pelvis, genitals and breasts (N40-N99)	140 (1.18)	0.92 (0.78- 1.09)
Skin conditions (L00-L99)	138 (1.15)	1.53 (1.29- 1.81)
Nervous system (G00-G99)	126 (1.40)	1.82 (1.51- 2.16)
Eye conditions (H00-H59)	120 (1.35)	0.62 (0.52- 0.74)
Blood disorders (D50-D89)	83 (2.84)	1.41 (1.12- 1.74)
Infectious diseases (A00-B99)	65 (1.20)	1.91 (1.48- 2.44)
Ear conditions (H60-H95)	10 (1.00)	0.82 (0.39- 1.50)
Congenital abnormalities (Q00-Q99)	6 (1.00)	0.78 (0.29- 1.70)

Table 7.3 displays the three-character ICD-10 diagnoses which accounted for at least 1% of all primary discharge diagnoses, and indicates that a few conditions accounted for multiple hospitalisation instances in the SMI cohort. The most common was chronic renal failure (N18) accounting for 18.5% of all admissions, but this was substantially less common as a cause of any admission (i.e. when repeat admissions were excluded), as the 1,592 hospitalisations occurred in only 22 individuals. Dental caries, other disorders of the urinary system, pneumonia, sickle cell disorders, type 2 diabetes mellitus, and chronic obstructive pulmonary disease were other common specific disorder diagnoses with SARs raised, in addition to non-specific diagnoses such as throat/chest pain, abdominal/pelvic pain, the Z04 code, and alcohol-related presentation. Apart from alcohol-related presentations (F10), all SARs remained statistically significant when repeat admissions were excluded. 'Other cataract' was the only one of these hospitalisation diagnoses which was less common in SMI than in the source population, and breast cancer the only diagnosis showing no statistically significant difference in admission rates. Considering ratios of all admissions to unique admissions in SMI (i.e. column 3 divided by column 5 in Table 7.3), the highest values were found for chronic renal failure (72.4), sickle cell disorders (7.6), breast cancer (5.7), and disorders related to alcohol use (2.0), with the remainder being below 1.5.

Table 7:3- Age- and gender-standardised admission ratios (SARs) for three-character ICD code diagnoses contributing to at least 1% of all hospitalisations in 2009-10 among individuals with SMI in Southeast London

		All hospitalisations		Excluding repeat hospitalisations	
ICD-10 code	Denoting clinical syndrome	Number (% of total)	Standardised admission ratio (95% CI)	Number	SAR
N18	Chronic renal failure	1,592 (18.5)	1.60 (1.53-1.68)	22	1.69 (1.06-2.56)
Z04	Examination and observation for other reasons	264 (3.1)	6.66 (5.88-7.51)	211	5.60 (4.87-6.41)
K02	Dental caries	222 (2.6)	2.55 (2.22-2.91)	159	2.24 (1.91-2.62)
N39	Other disorders of urinary system	173 (2.0)	2.65 (2.27-3.07)	137	2.55 (2.14-3.01)
R07	Pain in throat and chest	149 (1.7)	1.57 (1.33-1.85)	110	1.37 (1.12-1.65)
J18	Pneumonia, organism unspecified	144 (1.7)	3.33 (2.81-3.92)	126	3.26 (2.72-3.88)
C50	Malignant neoplasm of breast	120 (1.4)	0.87 (0.72-1.04)	21	0.95 (0.59-1.45)
R10	Abdominal and pelvic pain	115 (1.3)	1.56 (1.29-1.87)	84	1.27 (1.01-1.58)

		All hospitalisations		Excluding repeat hospitalisations	
ICD-10 code	Denoting clinical syndrome	Number (% of total)	Standardised admission ratio (95% CI)	Number	SAR
D57	Sickle-cell disorders	114 (1.3)	9.04 (7.45-10.9)	15	4.83 (2.70-7.97)
E11	Type 2 diabetes mellitus	101 (1.2)	4.46 (3.63-5.42)	73	4.28 (3.35-5..38)
F10	Mental and behavioural disorders due to use of alcohol	91 (1.1)	16.5 (13.3-20.2)	45	9.80 (0.15-13.1)
J44	Other chronic obstructive pulmonary disease	88 (1.0)	1.85 (1.48-2.28)	62	2.18 (1.67-2.79)
H26	Other cataract	87 (1.0)	0.49 (0.39-0.60)	63	0.50 (0.39-0.64)

7.6 Discussion

In this study, I sought to draw a profile of the recorded reasons for non-psychiatric hospital admissions in people with SMI. In terms of overall numbers of hospital admissions in SMI, the leading diagnostic groups represented urinary, digestive, neoplastic, respiratory and circulatory conditions. For all of these groups of conditions apart from circulatory disease, admission rates were significantly different to those from the general population with SARs higher than expected for urinary, digestive and respiratory conditions and lower than expected for neoplasms. Hospitalisations with non-specific diagnostic categories (ICD-10 R and Z codes) were also common and higher than in the general population as were those due to injuries, poisonings and other external causes, and those categorised as due to mental disorder. All of these observations held true when repeat admissions were excluded.

This investigation of non-psychiatric hospitalisations is, to my knowledge and based on literature reviewed, novel for SMI and was made possible because of data linkages negotiated and set up with CRIS over a long period. Other studies of hospitalisations have tended to focus either on mental healthcare specifically or all hospitalisations more generally (Jacobs et al. 2015; Kisely et al. 2015). In this study, considering individual diagnoses, clearly a number of these hospitalisations outside mental health services attracted a primary diagnosis of a mental or behavioural condition, with alcohol-related disorders accounting for around a third of instances and the most common specific three-digit diagnosis. The role of harmful drinking as a cause of hospitalisations requires further more detailed evaluation in SMI but is likely to encompass a range of factors including

admissions for intoxication or alcohol poisoning, inpatient detoxification episodes and potentially other health consequences assigned as alcohol-related in the recording of the primary diagnosis. Other hospitalisations due to direct impacts of underlying mental disorders are likely to include injuries, poisonings and other external causes arising secondary to self-harm and/or violence from others. Relationships between mental disorder and violence have been discussed extensively. A mental health policy paper that reviewed relevant literature concluded that SMI in itself does not cause an individual to be violent and instead it is the socio-demographic and socio-economic factors that lead to violence; the review also concluded that individuals with SMI are more likely to be victims of violence than perpetrators (Stuart H 2003). Substance misuse is clearly an important risk factor for perpetration and experience of violent behaviour both with and without co-occurring SMI (Stuart H 2003) and strategies to reduce this comorbidity may have potential benefits on a wide range of outcomes.

It is also interesting that hospitalisations assigned miscellaneous diagnostic codes occurred more often than expected in the SMI cohort. These might reflect clinically unexplainable symptoms or admissions precipitated by breakdowns in care arrangements and/or social support rather than secondary to a defined physical disorder. A recent project found that a small number of specialist mental health hospitals frequently used the ICD-10 code R69.x ("unknown diagnosis") as the primary diagnosis for people with a previous SMI diagnosis (White J, Gutacker N, Jacobs R 2014); however, I did not identify any literature to suggest that this was common practice within non-psychiatric settings. Clinically unexplainable symptoms underlying non-specific codes are possibly supported by the higher frequencies

of throat/chest pain and abdominal/pelvic pain. However, it should be borne in mind that pain in the throat and chest might have been caused by cardiac disease and it is well established that people with SMI have excess cardiovascular mortality (Leucht et al. 2007), suggesting a potential need for higher scrutiny of these symptoms, which might be complex to interpret when comorbid with mental disorders. Oesophageal reflux symptoms have been found to occur more frequently among people with a range of mental disorders, not just SMI, potentially indicating a reduced threshold for or distorted perception of symptoms (Avidan B, Sonnenberg A, Giblovich H 2001), although it is also important to bear in mind reflux associated with use of psychotropic drugs with sedative and/or anticholinergic actions, as well as an indirect effects of adverse lifestyle factors such as alcohol and substance misuse (Avidan B, Sonnenberg A, Giblovich H 2001).

Chronic renal failure predominated as a diagnosis for all admissions over the 2-year follow-up period, but this was accounted for by many admissions in a small number of individuals, most likely representing inpatient dialysis episodes and/or admissions for complications of renal disease. The fact that people with SMI had both higher than expected total numbers of admissions with chronic renal failure and were more likely to have at least one admission with this condition, indicates a higher prevalence and/or severity of renal disease in this population. Other studies have also shown that in people with chronic renal disease, co-occurring SMI can lead to an increased risk of experiencing another hospitalisation, particularly emergency admissions (McPherson et al. 2014). This may be accounted for by a higher burden of underlying risk factors such as hypertension and diabetes, although the role of medications such as lithium needs further evaluation, as does the possibility of

reverse causation through long-term renal problems leading to the development of psychotic illness (Tzeng et al. 2015; Fanton & Mcintyre 2011).

While the findings above are in line with published literature, the association I found with higher than expected sickle cell related hospitalisations is possibly the first of its kind. Sickle cell disease mainly affects people of African, Caribbean, Middle Eastern, Eastern Mediterranean and Asian origin and the high SARs in this cohort are concerning, given the ethnic profile of the comparator local general population (around 25% black and black British residents). I could not identify any supporting literature beyond a handful of case reviews of psychosis among individuals with sickle cell, thus indicating the need for further research. Another novel finding is that cataract-related hospitalisations were less common which might reflect a lack of access to health services, as literature did not highlight any mental disorder profiles which would confer protection. However, coding issues need to be considered further, as high number of "other cataracts" (ICD-10 code: H26) were identified. Studies elsewhere have identified that surgery for cataracts is more likely to happen as an inpatient for individuals with SMI than as a day case where discharge is on the same day (Dorning, H. Davies, A. Blunt 2015); therefore more rather than fewer cataract-related hospitalisations would have been expected.

There were also collections of associated diagnoses potentially indicative of social deprivation and lack of health service access. These included dental caries, chest infections and chronic obstructive pulmonary disease, with smoking also a potential underlying factor.

Dental caries, also known as tooth decay, is one of the two most common diseases that affect oral health, and people with SMI have been found to have over three times the odds of edentulousness than the general population (Kisely S, Quek LH, Pais J, Lalloo R, Johnson NW 2011). The reasons for higher susceptibility to oral disease include smoking, amotivation syndromes, worse oral hygiene, generalised anxiety or a specific fear of dental examinations and procedures, costs of dental care, difficulty in accessing healthcare facilities and side effects of psychotropic drugs such as xerostomia (Kisely S, Quek LH, Pais J, Lalloo R, Johnson NW 2011). However, one study revealed that poor oral hygiene and reduced access to dental care were most important in determining the sub-optimal oral health of individuals with SMI (Matevosyan 2010). Higher hospitalisations with pneumonia, organism unspecified, may also arise from self-neglect and impaired ability to recognise worsening of symptoms. In this respect, a lack of prompt detection and appropriate prevention of pneumonia have previously been found to increase the risk of poor outcomes including hospitalisation in schizophrenia (Yi-Hua C 2011). Poor treatment outcomes could also reflect misinterpretations of patients' complaints as psychosomatic, resulting in delayed recognition (Schoepf & Uppal 2014). Self-neglect and its impact is an important consideration; however, the term "self-neglect" has been loosely applied, and definitions have not yet been established or validated, despite some attempts to clarify its contribution to vulnerability (Lamkin J, Nguyen PT, Coverdale JH 2016). Non-specific coding may also be an indication that individuals with SMI may need more advocacy in order to gain equitable access to appropriate care and support, and this might involve assistance in articulating what would otherwise be classified as unexplained symptoms.

This study had several strengths, including the large SMI cohort and its naturalistic data on people receiving mental and physical healthcare in an urban/suburban neighbourhood. The outcome investigated was derived from a data source used in all national hospitals and was available for the whole of England. It should therefore have covered the vast majority of hospitalisations experienced both by the SMI cohort and the catchment population. The data linkage also provided the opportunity to separate out admissions only to non-psychiatric hospitals which, as mentioned, have received surprisingly little previous investigation. However, there are also limitations which need to be borne in mind. The study provides a descriptive analysis and not inferential statistics. The study was not able to distinguish balanced positive/negative effects at different points on the causal pathway; for example, breast cancer could conceivably be more common in SMI but inpatient healthcare less often sought/received. Furthermore, the study was restricted to assessing hospitalisations and did not attempt to investigate primary care outcomes or use of outpatient services. The study only investigated admissions to acute hospitals and there may of course have been individuals seen in mental health units for some physical illnesses. Finally, although the cohort is likely to be representative of people with SMI living in urban and suburban settings, there may be some unique features; for example, the prominence of sickle cell disease as a reason for hospitalisation may simply reflect the different ethnic compositions between the catchment general population and study cohort with SMI.

Non-psychiatric hospitalisation is an important outcome for a number of reasons. First, and most importantly, preventable admissions are likely to be occurring which are not in the best interest of the patient. Second, preventable hospitalisations cost money that could be

utilised differently for the benefit of the individual with SMI and the public purse. This study highlights that there is a range of disease groups (e.g. digestive and respiratory conditions), specific disorders (e.g. sickle cell disease and chronic renal failure) and associated scenarios (e.g. self-neglect and alcohol abuse) that are important for policy development. Future research could helpfully focus on identifying and evaluating appropriate models for delivering more effective care, particularly for the conditions where standardized admission ratios were higher than expected. However, investigation is also warranted where admissions were lower than that of the general population, to ensure that unmet needs are not being missed, given the recognized vulnerability of people living with SMI.

7.7 Chapter summary

The commonest five high-level categories of discharge diagnosis in the SMI cohort were urinary conditions, digestive conditions, unclassified symptoms, neoplasms, and respiratory conditions. SARs were raised for all of these apart from neoplasms which were lower than expected. Hospitalisations were also higher than expected for poisoning and external causes, injury, mental/behavioural disorders, endocrine/metabolic conditions, blood disorders, nervous disorders, skin disorders and infections, as well as non-specific ('Z') codes. The five commonest specific ICD-10 diagnoses were chronic renal failure (N18), a non-specific code (Z04), dental caries (K02), other disorders of the urinary system (N39) and 'pain in throat and chest' (R07), all of which were higher than expected.

8 Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method

8.1 Chapter outline

This chapter investigates the relationship between negative symptoms identified in the clinical records of a large sample of patients with schizophrenia using natural language processing and clinical outcomes.

The chapter is based on the following peer-reviewed paper-

Patel R*, Jayatilleke N*, Broadbent M, Chang CK, Foskett N, Gorrell G, Hayes RD, Jackson R, Johnston C, Shetty H, Roberts A, McGuire P, Stewart R. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. BMJ Open. 2015 Sep 7;5(9):e007619.

* contributed equally

Personal contributions to publication- Advice on study concept and design were sought from senior authors listed above. I had full access to all the data in the study and undertook the checks for data integrity and accuracy of analyses carried out by my co-author (Patel R). I took lead on drafting the manuscript and undertaking the critical revisions following co-

author feedback including referencing. My co-author (Patel R) led on the submission process to the publishing journal.

I have subsequently re-analysed all major findings with ethnicity as a covariate included in the dataset independently.

8.2 Research question

To identify negative symptoms in the clinical records of a large sample of patients with schizophrenia using natural language processing and assess their relationship with clinical outcomes.

8.3 Background

Negative symptoms, which include amotivation, a flattening of emotional responses, a reduction in speech and activity, and social withdrawal, contribute to much of the disability associated with schizophrenia (Foussias et al. 2011; Hunter & Barry 2012). These symptoms are also associated with poor psychosocial functioning and a reduced likelihood of remission (Jager et al. 2009; Ucok et al. 2011; Moller et al. 2000; Dominguez et al. 2010; McGurk et al. 2000; Kirkpatrick et al. 2006). The aetiology and pathophysiology of negative symptoms are unknown, and there are no effective treatments (Arango et al. 2013; Fusar-Poli P, Papanastasiou E 2014). With age, the presence of negative symptoms and its associated factors may change. There is suggestion that negative symptoms do not dominate the clinical picture in later life but do have association with positive symptom scores, lower cognitive scores and fewer confidantes (Cohen CI, et al 2013).

A number of rating scales have been developed to assess negative symptoms (Blanchard JJ, et al. 2011; Kirkpatrick B, et al. 2011; Kay et al. 1987). However, these are relatively detailed, require a trained rater, and are not routinely applied in clinical practice. As a result, current knowledge of negative symptoms is derived from studies in relatively small samples of patients, who may have been selected for inclusion because they had particularly severe symptoms. The findings from these samples may not therefore be representative of negative symptoms in the overall population of patients with schizophrenia.

Clinical information is increasingly recorded electronically, facilitating access of rich clinical data, including presence or absence of symptoms (Patel R, Lloyd T 2015), from routine medical records. In the present study, a novel information extraction tool to identify negative symptomatology in a large body of electronic records collected from individuals with schizophrenia was used (Gorrell G, Jackson R, Roberts A 2013; Cunningham H, Tablan V, Roberts A 2013; Patel R, Jayatilleke N 2014). The relationships between negative symptoms and clinical outcomes were then examined. The hypothesis was tested that negative symptoms are associated with poor clinical outcome in patients with schizophrenia, as indexed by the frequency and duration of hospital admissions. A supplementary analysis was performed with age stratification on the association between negative symptoms and clinical outcomes.

8.4 Methods

8.4.1 Inclusion criteria, outcomes, analysis

Three samples were identified for analysis and extracted from CRIS:

I. Sample A (n=7678): patients with schizophrenia (International Classification of Diseases (ICD)-10 F20.XX) aged 16 years and over who had used SLaM services during 2011. This sample was used to investigate: (1) the relationship between negative symptoms, documented at any point in the electronic health record, and demographic and other clinical measures (described below); (2) the relationship between negative symptoms documented prior to 1 January 2011 and the risk of hospital admission during 2011. This year was chosen for analysis because it maximised the duration of time over which text would be available for measurement development, while allowing at least 12 months follow-up in all instances.

II. Sample B (n=1612): the subset of patients from sample A who had been discharged from SLaM inpatient care during 2011. This sample was used to investigate the relationship between negative symptoms documented prior to 2011 and the risk of readmission in the 12 months following discharge.

III. Sample C (n=1609): the subset of patients from sample A who received SLaM inpatient care during 2011. This sample was used to investigate the relationship between negative symptoms documented prior to 2011 and the length of the first hospital admission during 2011.

8.4.2 Measurement development

Natural language processing (NLP) information extraction allows structured information to be obtained from unstructured text records. NLP to detect statements in the correspondence fields of clinical records to determine references to pre-specified negative symptoms was carried out. Full details of the NLP method are described in a previous paper (Jackson et al. 2013) and are outlined in chapter 5. In summary, a putative training data set was selected which contained broad dictionary terms relevant to the negative symptoms of interest (described below). A detailed review of the training data set was undertaken by two psychiatrists to identify and annotate key phrases within the records that were either relevant or irrelevant for keywords related to each symptom. Inter-rater reliability was tested between the two annotators resulting in percentage agreement of 93.0% (Cohen's κ 0.85). This training data set annotated by professionals was the gold standard used to construct an application (CRIS Negative Symptoms Scale, CRIS-NSS) using a hybrid classification model consisting of a support vector machine (SVM) learning algorithm (Li et al. 2009) and rule-based text matching, using the Generalised Architecture for Text Engineering (GATE) software package (Cunningham et al 2013). Following the same approach as the development of the training dataset, a gold standard data set was developed (Jackson et al. 2017). The SVM algorithm was applied using a 'bag-of-words' approach to take into account the context of negative symptoms within the sentence in which they were documented, thereby allowing ascertainment of negative symptoms experienced specifically by the patient as well as distinguishing between positive instances and negated instances (Jackson et al. 2013). Once developed, CRIS-NSS was subsequently

used to determine the presence of negative symptoms within the clinical sample. The accuracy of CRIS-NSS was evaluated using precision and recall statistics which were generated through internal fivefold cross-validation (Li et al. 2009): precision, representing the proportion of text instances identified by the tool which were found to be correct in terms of identifying the negative symptom of interest (equivalent to positive predictive value); and recall, measuring the proportion of text instances recording a given negative symptoms which were correctly identified as such by the tool (equivalent to sensitivity).

All symptoms were extracted in the same way using similar Natural Language Processing steps with applications developed for 10 items: poor motivation, blunted or flattened affect, poor eye contact, emotional withdrawal, poor rapport, social withdrawal, poverty of speech, mutism, apathy and concrete thinking. Symptom definitions that were developed from initial annotation guidelines are shown in the appendix 1. Each of these symptoms was defined as a binary variable on the basis of being present at any point in the record within the defined time period, and a composite scale (range 0–10) was constructed by summing these variables, followed by Cronbach α score calculation (a measure of inter-correlation between individual scale items) to estimate its internal consistency. A threshold score of at least 2 (i.e. two or more negative symptoms documented) was applied a priori to determine the presence or absence of negative symptoms for analysis as a binary variable, as well as treating the scale score as an ordinal variable. Further technical details of the criteria for ascertaining the negative symptoms in the CRIS-NSS application are described elsewhere (Jackson et al. 2013).

8.4.3 Clinical outcome measures and covariates

The following clinical and demographic variables were obtained as covariates from the data set: date of birth, gender and ethnicity were extracted in addition to employment, marital status and ethnicity nearest to 1st January 2011. Date of birth was used to calculate age on 1st January 2011 and managed as a continuous variable. Ethnicity classifications were- "White", "Black", "Asian", "mixed, unknown and others". Marital status was classified as being "single", "in a relationship (cohabiting, married or civil partnership)", "separated", "not disclosed or unknown". Employment was classified as "employed", "unemployed" or "unknown". Using structured data derived from the Health of the Nation Outcome Scale (HoNOS)(Wing JK, Beevor AS 1998), routinely completed in SLaM patients, the following subscales (scored 0–4) were used as covariates: activities of daily living (ADL) impairment, problems with relationships (social impairment), presence of hallucinations or delusions (a measure of positive symptoms) and depressive symptoms. For all of these HoNOS subscales, binary variables were defined on the basis of a score of 2 or more indicating the presence of each construct at levels judged to be clinically significant. In cases with multiple data points, all covariates were defined as those recorded closest to 1 January 2011. The date of admission and date of discharge were used to calculate duration of stay in days. Readmission was considered as a subsequent admission within the 12 months following discharge from index admission.

8.4.4 Statistical analysis

STATA (V.11) software was used. Estimates of prevalence of negative symptoms by demographic factors were obtained as the proportion of patients within each group with two or more negative symptoms. After describing the distribution of negative symptoms and the psychometric properties of the CRIS-NSS, further analyses were performed to investigate the associations between the clinical outcomes described above and (1) the presence of negative symptoms, using binary logistic regression; and (2) CRIS-NSS scores, using ordinal logistic regression. Reference groups for categorical variables were generally defined as the most prevalent category, apart from age group where 20-29 years group, the youngest group of sufficient size was assigned as the reference. Associations between negative symptomatology and hospital admission and readmission were analysed using logistic regression, while those with length of inpatient stay were analysed using linear regression—again, estimating associations with both the binary and ordinal CRIS-NSS exposure. For the analyses with hospitalisation outcomes in/following 2011, CRIS-NSS was generated restricting information extraction to electronic health records prior to 2011. Where data were missing on individual covariates (in 2362 participants), this was indicated in the regression models as a separate category, supplemented by sensitivity analyses performed on the sample with complete data on all covariates to check the consistency of findings (appendix 4- table 2). For the supplementary analysis, the previous analyses were repeated within the subgroups of those aged under the age of 40 years and those over the age of 40 years and including an interaction term of age under or over 40 and binary CRIS-NSS exposure. Finally, secondary analyses were undertaken to investigate and compare the

relationships of individual CRIS-NSS symptoms with risk of readmission and length of stay using binary logistic and linear regression, respectively.

8.5 Results

8.5.1 Performance of CRIS-NSS

Table 8.1 illustrates results from fivefold cross-validation of the CRIS-NSS tool. Precision coefficients ranged between 0.80 (poverty of speech) and 0.99 (mute) and recall between 0.62 (poor motivation) and 0.97 (apathy). For the composite 10-point scale, the Cronbach α value was 0.78 indicating a good level of internal consistency.

Table 8:1- Performance of Clinical Record Interactive Search Negative Symptoms Scale (CRIS-NSS) information extraction applications ascertaining individual symptom domains (n=7678)

Symptom	Precision/ recall	Prevalence (%) in patients with schizophrenia receiving care during 2011
Poor motivation	0.87/0.62	30.5
Blunted or flattened affect	0.93/0.83	27.4
Poor eye contact	0.95/0.79	26.0
Emotional withdrawal	0.85/0.74	23.5
Poor rapport	0.91/0.77	16.3
Social withdrawal	0.94/0.96	12.7
Poverty of speech	0.80/0.73	12.4
Mute	0.99/0.94	8.1
Apathy	0.88/0.97	7.7
Concrete thinking	0.91/0.72	5.7

8.5.2 Prevalence and distribution of negative symptoms

Of the 7678 patients with schizophrenia, 3149 (41.0%) had at least two negative symptoms documented. Table 8.1 displays prevalence for each of the symptoms classified by the tool. The most frequently recorded symptoms were poor motivation (30.5%), blunted or flattened affect (27.4%), poor eye contact (26.0%) and emotional withdrawal (23.5%). The prevalence by number of symptoms were as follows: one symptom 14.6%, two symptoms 12.7%, three symptoms 9.3%, four symptoms 6.4%, five symptoms 5.0%, six or more symptoms 7.6%.

Table 8.2 shows prevalence of two or more negative symptoms and binary logistic regression analyses for factor associated with this category. The prevalence of two or more symptoms was highest for age 20-29 years group (52.0%); 45.3% of males had two or symptoms compared to 34.7% females; 44.6% of individuals who were single had two or more symptoms compared to 31.6% of individuals married or cohabiting and 33.4% of individuals who divorced or separated had two or more symptoms. Patients who were employed were less likely to have negative symptoms (39.6%) compared with those unemployed (47.9%), and prevalence were higher in Black (48.2%) and Asian (38.7%) compared to White (31.0%) ethnic groups. When adjusted for other factors, two or more negative symptoms were also significantly associated with ADL impairment (p -value<.001) but not social impairment (p-value= 0.24). After adjustments, presence of delusions/ hallucinations (p-value=0.001) and depression (<0.001) were significantly associated with having two or more symptoms.

Table 8:2- Binary logistic regression analysis of factors associated with negative symptoms in patients with schizophrenia (n=7678)

Factor	Group	Number in sample	Prevalence of two or more negative symptoms (%)	Association with two or more negative symptoms: Unadjusted OR (95% CI)	P value	Association with two or more negative symptoms: Adjusted model OR (95% CI) (n=7676)*	P value
Age (years)	16–19	203	27.6	0.35 (0.25 to 0.49)	<0.001	0.52 (0.37 to 0.73)	<0.001
	20–29	1337	52.0	Reference		Reference	
	30–39	1775	47.0	0.82 (0.71 to 0.94)	0.006	1.04 (0.89 to 1.21)	0.612
	40–49	1983	42.6	0.69 (0.60 to 0.79)	<0.001	0.96 (0.82 to 1.11)	0.615
	50–59	1137	37.2	0.55 (0.47 to 0.64)	<0.001	0.93 (0.83 to 1.10)	0.397
	60–69	654	29.1	0.38 (0.31 to 0.46)	<0.001	0.68 (0.55 to 0.84)	<0.001

Factor	Group	Number in sample	Prevalence of two or more negative symptoms (%)	Association with two or more negative symptoms: Unadjusted OR (95% CI)	P value	Association with two or more negative symptoms: Adjusted model OR (95% CI) (n=7676)*	P value
	70+	589	18.0	0.20 (0.16 to 0.26)	<0.001	0.45 (0.36 to 0.57)	<0.001
Gender	Male	4592	45.3	Reference		Reference	
	Female	3084	34.7	0.64 (0.59 to 0.71)	<0.001	0.75 (0.68 to 0.83)	<0.001
Marital status (most recent)	Single	5795	44.6	Reference		Reference	
	Married/cohabiting	785	31.6	0.57 (0.49 to 0.67)	<0.001	0.75 (0.64 to 0.89)	0.001
	Divorced/separated	776	33.4	0.62 (0.53 to 0.73)	<0.001	0.76 (0.65 to 0.90)	0.001
	Widowed	208	21.2	0.33 (0.24 to 0.47)	<0.001	0.75 (0.54 to 1.03)	0.079

Factor	Group	Number in sample	Prevalence of two or more negative symptoms (%)	Association with two or more negative symptoms: Unadjusted OR (95% CI)	P value	Association with two or more negative symptoms: Adjusted model OR (95% CI) (n=7676)*	P value
Employment (most recent)	Unemployed	4956	47.9	Reference		Reference	
	Employed	341	39.6	0.71 (0.57 to 0.89)	0.003	0.72 (0.58 to 0.91)	0.006
	In education	311	39.6	0.71 (0.56 to 0.90)	0.004	0.84 (0.65 to 1.07)	0.152
	Retired	7	14.3	0.18 (0.02 to 1.51)	0.114	0.68 (0.14 to 3.19)	0.627
Ethnicity	White (British, Irish, any white background)	3396	31.0	Reference		Reference	
	Black (African, Caribbean, any Black background)	3321	48.2	2.07 (1.88- 2.29)		1.80 (1.61- 2.00)	<0.001

Factor	Group	Number in sample	Prevalence of two or more negative symptoms (%)	Association with two or more negative symptoms: Unadjusted OR (95% CI)	P value	Association with two or more negative symptoms: Adjusted model OR (95% CI) (n=7676)*	P value
	Asian (Indian, Chinese, any Asian background)	498	38.7	1.40 (1.15- 1.71)		1.55 (1.26- 1.90)	<0.001
	Mixed/ unknown/ missing	593	33.6	1.12 (0.93- 1.35)		1.04 (0.86- 1.27)	0.648
Activities of Daily Living impairment	Absent	4700	41.9	Reference		Reference	
	Present	2283	46.3	1.20 (1.08 to 1.32)	<0.001	1.33 (1.20 to 1.52)	<0.001
Social impairment	Absent	4432	42.7	Reference		Reference	
	Present	2533	44.4	1.07 (0.97 to 1.18)	0.158	0.93 (0.83 to 1.04)	0.217

Factor	Group	Number in sample	Prevalence of two or more negative symptoms (%)	Association with two or more negative symptoms: Unadjusted OR (95% CI)	P value	Association with two or more negative symptoms: Adjusted model OR (95% CI) (n=7676)*	P value
Delusions/hallucinations	Absent	3904	41.9	Reference		Reference	
	Present	3077	45.0	1.14 (1.03 to 1.25)	0.009	1.30 (1.17 to 1.44)	<0.001
Depression	Absent	4976	45.2	Reference		Reference	
	Present	2014	38.8	0.77 (0.69 to 0.85)	<0.001	0.66 (0.59 to 0.74)	<0.001

*Results adjusted for all the factors reported in this table; two cases with no recorded data on gender were dropped

8.5.3 Hospital admission, length of stay and readmission

Figure 8.1 summarises the association of negative symptoms recorded prior to 2011 with mental health admission and readmission in 2011. The percentage hospital admissions were highest for individuals who six or more symptoms (36.8%) while percentage hospital readmission was highest for individuals with five symptoms (47.1%). Figure 8.2 summarises length of hospitalisation for inpatients during 2011. The median duration of stay was lowest for individuals who had no symptoms (30days) while highest for individuals with six or more symptoms (75 days). Further details including interquartile ranges are available in appendix 4 table 6. Table 8.3 shows the findings from the logistic and linear regression analyses. confirmed that negative symptoms were associated with a higher likelihood of admission, readmission and a longer duration of hospitalisation. Specifically, after full adjustment (Table 8.3, model 3), patients with two or more negative symptoms before 2011 had a 24% greater likelihood of admission during 2011. Moreover, each of their admissions was, on average, an extra 21 days in duration, and when they were discharged, they had a 58% higher risk of readmission within 12 months. All of these associations remained independent and largely unaltered following adjustment for intensity of delusions/hallucinations among other covariates. Further analysis comparing patients aged under and over 40 years showed that the effects of negative symptoms on inpatient admission were broadly similar for both groups but with a slight increase in risk of readmission and reduced duration of admission in relation to negative symptoms for those under 40 compared with those over 40. However, the age × negative symptoms interaction term remained a non-significant factor ($p>0.05$) for all models.

Figure 8-1- Percentage of patients admitted to hospital (sample A) or readmitted to hospital following discharge (sample B) in 2011 by number of negative symptoms

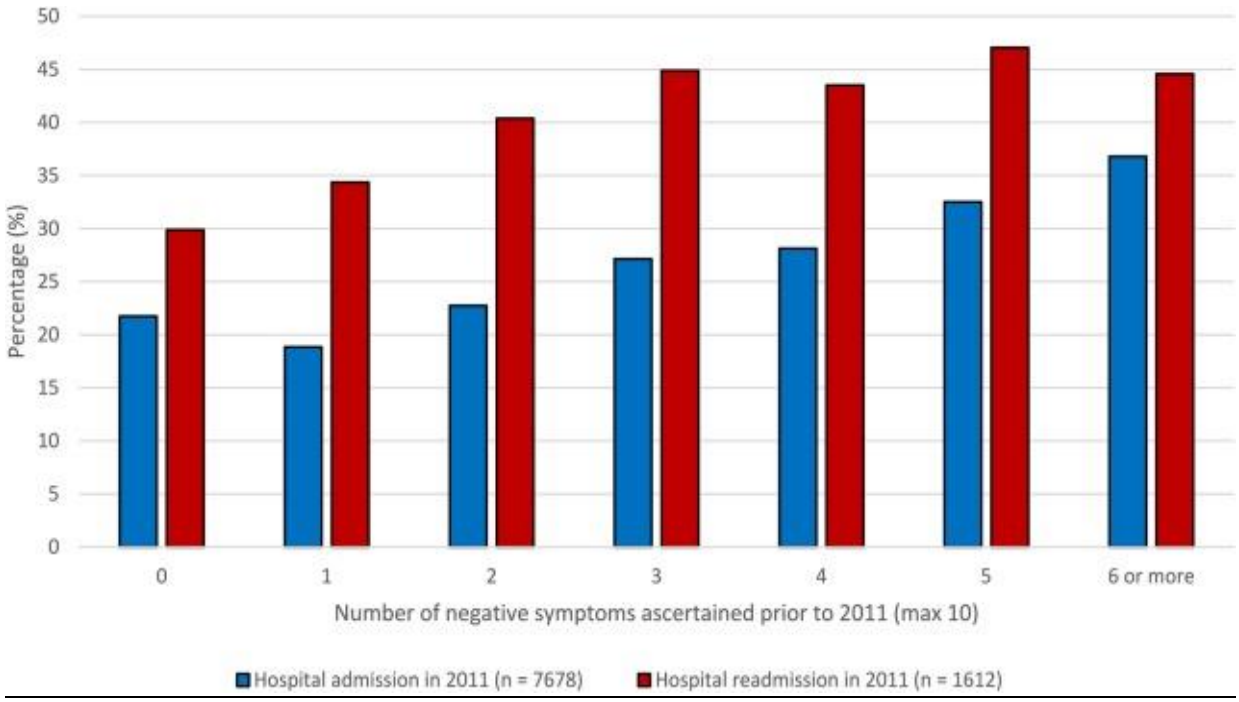


Figure 8-2- Median duration of admission among mental health inpatients with schizophrenia (sample C) in 2011 by number of negative symptoms (n=1609)

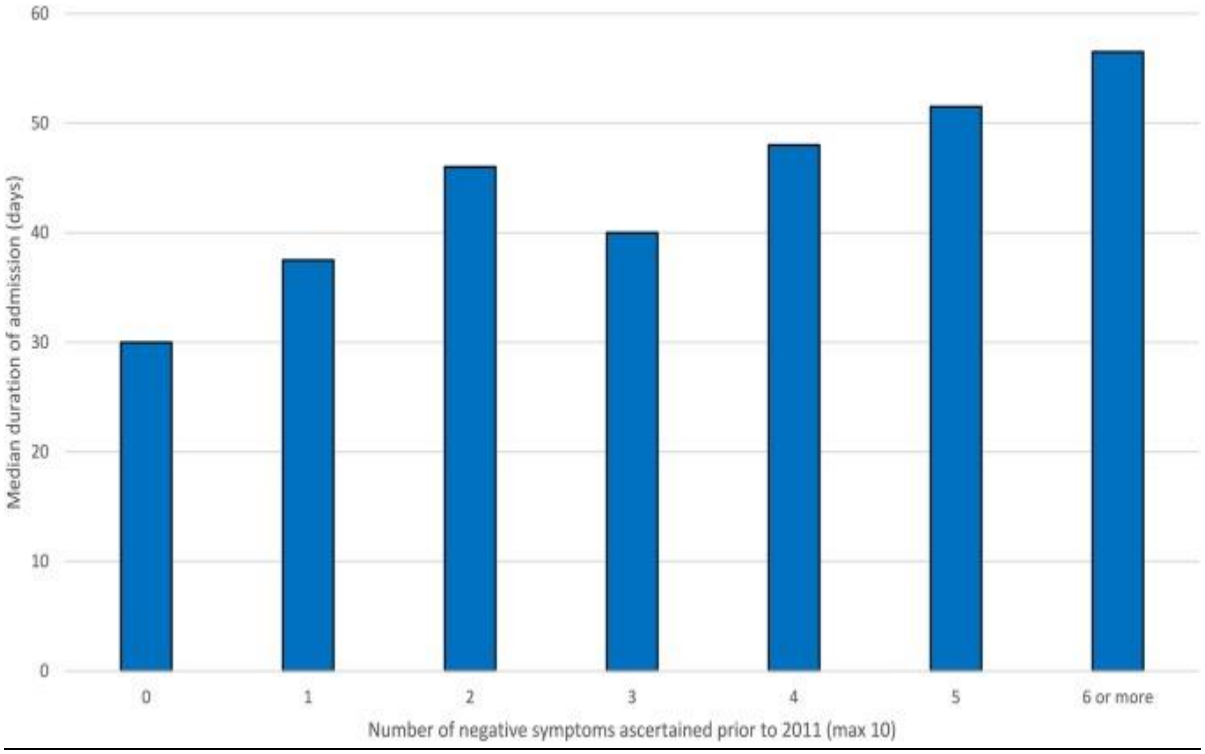


Table 8:3- Association between number of negative symptoms ascertained prior to 2011 and mental health hospital admission, readmission and duration of admission in 2011

	Inpatient admission (OR, 95% CI; n=7678)*	Readmission within 12 months of inpatient admission (OR, 95% CI; n=1612)*	Duration of inpatient admission (days; β-coefficient, 95% CI; n=1609)†
	Sample A	Sample B	Sample C
Associations with 2 or more negative symptoms (binary variable)			
Unadjusted	1.47 (1.32 to 1.63)	1.73 (1.41 to 2.12)	23.9 (11.2 to 36.7)
1. Age and sex	1.37 (1.23 to 1.53)	1.70 (1.38 to 2.09)	24.1 (11.3 to 36.9)
2. Model 1 plus marital status, employment and ethnicity	1.81 (1.61 to 2.02)	1.54 (1.22 to 1.96)	20.6 (7.5 to 33.7)

	Inpatient admission (OR, 95% CI; n=7678)*	Readmission within 12 months of inpatient admission (OR, 95% CI; n=1612)*	Duration of inpatient admission (days; β-coefficient, 95% CI; n=1609)†
	Sample A	Sample B	Sample C
3. Model 2 plus delusions/hallucinations, and depression	1.77 (1.58 to 1.99)	1.54 (1.23 to 1.93)	21.0 (8.0 to 34.1)
Associations with incremental number of negative symptoms (10-point scale ordinal variable)‡			
Unadjusted	1.12 (1.09 to 1.15)	1.12 (1.07 to 1.17)	6.5 (3.5 to 9.4)
1. Age and sex	1.09 (1.06 to 1.12)	1.11 (1.06 to 1.16)	6.3 (3.3 to 9.2)
2. Model 1 plus marital status, employment and ethnicity	1.16 (1.13 to 1.19)	1.01 (1.00 to 1.02)	5.6 (2.6 to 8.7)
3. Model 2 plus delusions/hallucinations, and depression	1.16 (1.13 to 1.19)	1.01 (1.00 to 1.01)	5.8 (2.8 to 8.9)

*Logistic regression. †Linear regression ‡ORs and β-coefficients are per one unit increase on the 10-point scale.

Table 8.4 shows the logistic and linear regression analyses that were performed to examine the relationship between individual negative symptoms and the frequency and duration of admission. Poor eye contact-OR 1.49 (95% CI 1.46- 1.81), poor motivation-OR 1.29 (95% CI 1.03- 1.60), emotional withdrawal OR 1.49 (95% CI 1.19- 1.87) poor rapport OR 1.50 (95% CI 1.16- 1.87) and mutism 1.56 (95% CI 1.50- 2.12) were associated with increased risk of readmission. Apathy was associated with increased duration of admission (p-value 0.032) but not with readmission risk (p-value 0.692). Emotional withdrawal and mutism were associated with both the risk of readmission and the duration of admission.

Table 8:4- Associations between individual Clinical Record Interactive Search Negative Symptoms Scale (CRIS-NSS) components and readmission risk/duration of admission in 2011

	Readmission risk- Sample B (binary logistic regression) (n=1612)				Duration of admission- Sample C (linear regression) (n=1590)			
	Unadjusted		Adjusted*		Unadjusted		Adjusted*	
	OR (95% CI)	p Value	OR (95% CI)	p Value	β-coefficient (95% CI)	p Value	β-coefficient (95% CI)	p Value
Poor motivation	1.40 (1.13 to 1.74)	0.002	1.29 (1.03 to 1.60)	0.026	23.0 (9.1 to 36.9)	0.001	19.1 (5.0 to 33.2)	0.008
Blunted or flattened affect	1.34 (1.08 to 1.65)	0.007	1.18 (0.95 to 1.48)	0.129	12.8 (−1.2 to 26.8)	0.073	8.3 (−5.7 to 22.4)	0.242
Poor eye contact	1.60 (1.30 to 1.98)	<0.001	1.48 (1.19 to 1.83)	<0.001	18.0 (4.2 to 31.8)	0.011	14.8 (0.9 to 28.6)	0.036
Emotional withdrawal	1.62 (1.30 to 2.02)	<0.001	1.49 (1.19 to 1.87)	0.001	32.5 (18.1 to 46.9)	<0.001	30.0 (15.6 to 44.4)	<0.001
Poor rapport	1.63 (1.29 to 2.06)	<0.001	1.50 (1.18 to 1.90)	0.001	23.1 (7.5 to 38.6)	0.004	21.1 (5.5 to 36.6)	0.008
Social withdrawal	1.16 (0.88 to 1.54)	0.291	1.02 (0.76 to 1.36)	0.911	16.4 (−2.9 to 35.7)	0.095	9.2 (−10.1 to 28.6)	0.349

	Readmission risk- Sample B (binary logistic regression) (n=1612)				Duration of admission- Sample C (linear regression) (n=1590)			
	Unadjusted		Adjusted*		Unadjusted		Adjusted*	
Poverty of speech	1.30 (0.98 to 1.70)	0.064	1.12 (0.85 to 1.49)	0.416	13.2 (−5.8 to 32.2)	0.173	8.5 (−10.5 to 27.5)	0.379
Mute	1.71 (1.27 to 2.30)	<0.001	1.56 (1.15 to 2.12)	0.004	28.5 (7.9 to 49.1)	0.007	29.2 (8.6 to 49.7)	0.005
Apathy	1.02 (0.71 to 1.47)	0.914	0.93 (0.64 to 1.35)	0.692	32.5 (6.7 to 58.2)	0.013	27.4 (1.8 to 53.1)	0.036

*Adjusted for age, sex, marital status, employment status, ethnicity, presence of hallucinations/delusions and depression.

After describing the distribution of negative symptoms and the psychometric properties of the CRIS-NSS, further analyses were performed to investigate the associations between the clinical outcomes described above and CRIS-NSS scores, using ordinal logistic regression. The findings are shown in appendix 4 table 1. For the analyses with hospitalisation outcomes in/following 2011, CRIS-NSS was generated restricting information extraction to electronic health records prior to 2011. Where data were missing on individual covariates (in 2362 participants), this was indicated in the regression models as a separate category, supplemented by sensitivity analyses performed on the sample with complete data on all covariates to check the consistency of findings – findings are shown in appendix 4 table 2. For the supplementary analysis, the previous analyses were repeated within the subgroups of those aged under the age of 40 years and those over the age of 40 years and including an interaction term of age under or over 40 and binary CRIS-NSS exposure- findings are shown in appendix 4 table 3.

8.6 Discussion

Using a SVM learning method with an NLP tool, it was possible to successfully to extract data on negative symptoms from the electronic mental health records of a large clinical sample of patients with schizophrenia. This approach did not require any specialised training or extra clinical assessments, and was able to generate a scale with robust construct and predictive validity from data recorded as part of routine clinical care. The results suggest that negative symptoms are documented in the electronic health records of a sizeable proportion of patients with schizophrenia, particularly in those who are relatively young, male and not in a relationship, consistent with findings from studies that assessed negative symptoms using quite different methods (Bobes et al. 2010; Cohen et al. 2013). These findings were based on the unprompted documentation of negative symptoms in the context of routine clinical care by staff who were not specifically trained in their assessment. Previous findings have usually been based on systematic ratings by a researcher using a dedicated rating scale. Negative symptoms are relatively difficult to detect and assess (Blanchard JJ et al 2011), and may be less frequently documented than positive symptoms, such as delusions and hallucinations, because they are less clinically obvious. In addition, mental health services in the UK are often orientated towards the management of acute crises, and hence the treatment of positive symptoms (Jacobs R 2011). It is thus possible that the figures for the prevalence and the severity of negative symptoms derived from this approach are lower than would have been obtained from a trained assessor using a standardised instrument. In addition, this method may be more likely to identify the types of negative symptoms (e.g. poverty of speech) whose detection does not require specialised training. A substantial proportion (41%) of the sample had at least two negative symptoms.

Although negative symptoms were defined and assessed in different ways to previous studies, this figure is comparable to that described in other samples of patients with schizophrenia ((Jager et al. 2009): 44%; (Bobes et al. 2010): 58%; (Cohen et al. 2013): 40%). Taken together, these findings suggest that negative symptoms are a relatively common feature of schizophrenia, rather than being limited to a subgroup of patients with a chronic, unremitting illness (Bagney et al. 2013). A clear association between negative symptoms and poor clinical outcomes was found, as indexed by impairments in daily living, increased risk of admission, increased duration of admission and increased risk of readmission. Hospital admissions are the main drivers of cost in the care of patients with schizophrenia (Department of Health. 2013), but have traditionally been linked to the severity of positive psychotic symptoms (Olfson M, Ascher-Svanum H 2011). This study indicates that negative symptoms are an equally important factor, and suggest that a greater emphasis on assessing and treating these features of schizophrenia may have significant health economic benefits. However, as these findings are drawn from observational data, it would be necessary to perform interventional clinical studies to determine whether an effective treatment for negative symptoms would lead to better clinical outcomes.

A key strength of this study was the large size of the patient sample, and that it was representative of the overall clinical population of people with schizophrenia in a defined geographic area. Previous studies of negative symptoms have usually involved smaller patient samples that were recruited to a research project (Jager et al. 2009; Bobes et al. 2010; Cohen et al. 2013). Focusing the information extraction process on text from correspondence maximises the generalisability of this approach, as letters to primary care

physicians (which accounted for a large portion of the correspondence text) are unlikely to vary substantially between mental health services with respect to the language used to describe the symptoms of interest.

In this analysis there were several limitations. Inpatient care outcomes were ascertained from CRIS, therefore only capture admissions to SLaM. Using an alternative like Hospital Episode Statistics data would have enabled a more comprehensive understanding of admissions. The outcomes chosen also do not include other clinical encounters such as in outpatient care. Adjustments were made for key demographic factors and the large sample size enabled multivariable analyses. However, there were several variables that were not included in the analyses including medication use which is a limitation. For the exposure, the patient's entire record rather than discrete periods of illness was searched, and it was not possible to delineate the timing or duration of individual negative symptoms, or whether they were primary (i.e. a direct consequence of illness) or secondary (e.g. side effects of treatment) as these measures were not routinely documented in electronic health records. Although the association of negative symptoms in clinical documents prior to 1 January 2011 with outcomes occurring after 1 January 2011 was investigated (to ensure that negative symptoms were always ascertained prior to outcomes), if negative symptoms were identified prior to 1 January 2011, it was not possible to ascertain when they occurred prior to this date, or their temporal relationships to subsequent clinical outcomes. The findings were thus derived from assessments made over a period that was not standardised, but was generally relatively long. This also introduces confounding by indication as unwell individuals may be more likely to be clinically reviewed more frequently and thereby have

their symptoms recorded more thoroughly. In contrast, most assessments of negative symptoms in the literature are derived from a single cross-sectional measurement (Gilbert et al. 2000; Chang et al. 2011). Measurement bias is likely to be low at the point of clinical care as clinical staff entering information during clinical care were not aware of study. At the stage of symptom applications development where training sets were developed and the clinical team were annotating individual symptoms, there is possibility of introducing a misclassification bias. This was handled with at least two team members conducting independent annotations and inter-rater agreement measured for each item. Also, while the annotator was aware the case notes were for an individual with SMI, they did not know the exact diagnosis or any other information other than the sentences on either side of the annotating word or phrase.

A further limitation was the extent to which individual negative symptoms could be considered as having equal weight in a composite score. Weighting the 10 negative symptom applications equally resulted in a composite score (from 0 to 10) with a reasonable degree of internal consistency, as demonstrated by a Cronbach α value of 0.78. However, analysing the association of each negative symptom with clinical outcomes revealed varying degrees of association with poor clinical outcomes for different negative symptoms. Future studies are necessary to examine the propensity for different negative symptoms to co-occur in individual patients and the extent to which different clusters of symptoms are associated with clinical outcomes, particularly in the light of previous research which suggests that negative symptoms segregate into two subdomains relating to amotivation and reduced emotional expression (Kimhy et al. 2006). The application of NLP

to clinical records is unlikely to identify negative symptoms as accurately as a direct assessment using a specialised psychopathological rating scale. However, automated tools could be used to screen individuals and identify those with negative symptoms who would then benefit from comprehensive assessment using a standardised instrument. In this way, automated methods could be used to complement standardised instruments. Automated information extraction tools could also be developed to identify other clinical parameters from electronic health records in order to support real-time clinical decision-making. These possibilities could be explored in future research. In summary, the findings suggest that negative symptoms can be identified in clinical records using automated methods, are common in patients with schizophrenia and are associated with poor clinical outcomes. The findings highlight the potential of automated information extraction tools in mental health research and clinical practice, and the importance of developing effective treatments for negative symptoms.

8.7 Summary of findings from chapter

Ten different negative symptoms were ascertained using NLP algorithms, with precision statistics above 0.80. Of patients with schizophrenia diagnoses, 41% had two or more negative symptoms recorded prior to the observation period. Negative symptoms were associated with younger age, male gender and single marital status, and with increased likelihood of hospital admission (OR 1.24, 95% CI 1.10 to 1.39), longer duration of admission (β-coefficient 20.5 days, 7.6–33.5), and increased likelihood of readmission following discharge (OR 1.58, 1.28 to 1.95). Negative symptoms were common and associated with adverse clinical outcomes, consistent with evidence that these symptoms account for much of the disability associated with schizophrenia. NLP provides a means of conducting research in large representative samples of patients, using data recorded during routine clinical practice.

9 Symptoms identified in electronic medical records for individuals with severe mental illness by natural language processing in relation to general mortality and hospitalisation

9.1 Chapter outline

This chapter describes an investigation of mental health symptom profiles in the electronic clinical records of a large cohort of patients with schizophrenia, schizoaffective disorder, or bipolar disorder in relation to all-cause mortality and measures of hospitalisation to non-mental health facilities.

The paper reporting this chapter is currently under peer review

Jayatilleke N, Chang CK, Stewart R. Symptoms identified in electronic medical records for individuals with severe mental illness by natural language processing in relation to general mortality and hospitalisation [In submission].

Personal contributions to publication- Using concepts discussed with supervisors, I designed the study approach. I had full access to all the data in the study and took responsibility for data integrity, accuracy of the data analysis, interpretation of data, drafting and critical revisions of the manuscript.

9.2 Research question

This study aimed to describe the associations with mortality and hospitalisation for each of six symptom dimensions (positive, negative, manic, disorganisation, catatonic and depressive) extracted from the clinical records of individuals with SMI.

9.3 Background

Symptom profiles, in a limited capacity, have been used to investigate adverse outcomes for individuals with SMI. To date, incorporation of symptoms in analyses investigating adverse outcomes have included the use of scores on global scales such the Health of the Nation Outcome scale as described in the study by Hayes et al., 2012. The use of specific symptoms has also been used, for example in the study reported by Fiedorowicz et al., 2009, which focused on manic and hypomanic symptoms in patients with bipolar disorder. It has long been recognised that people with SMI do not live as long as those without mental illness (Chang CK et al. 2011; Lawrence D et al 2013; Hoang U et al 2011). The majority of this excess mortality has been linked to physical illness with important individual lifestyle choices, side effects of psychotropic treatment and disparities in health care access, utilisation and provision contributing to these deaths as well as other poor physical health outcomes (De Hert et al. 2011). Higher suicide rates are also clearly important, but are by no means the predominating reason for loss of life expectancy in SMI (Druss BG, 2011). In its report as part of the World Psychiatric Association Action Plan 2008-2011, the authors note that this excess morbidity and mortality is largely due to modifiable lifestyle risk factors, but also that the screening and assessment of physical health disorders remains poor for people with SMI, even in developed countries (De Hert et al. 2011).The types of symptoms experienced may influence these outcomes.

Colleagues have highlighted that, currently, there are no internationally agreed standards to measure how many people with SMI need treatment (Thornicroft G 2013). As a result,

specific issues pertinent to individuals with SMI remain unresolved, including knowledge of the mental and physical outcomes of treatment or the absence of treatment, and there needs to be further work to understand how inequalities can be addressed. Symptoms expressed by a patient in the course of their illness, guide clinician decisions on its management including suitability for specific interventions and treatment modality, and symptom profiles provide more detailed information that just a diagnosis (Adam 2013; Chmielewski et al. 2014). Information on specific symptoms is typically recorded in unstructured parts of an individual's electronic health record (Hardoon et al. 2013) although NLP can be used to process large quantities of text and return structured information (Antolík 2005; Nikiforou et al. 2013; Chapman et al. 2011). The analysis described here utilised a unique opportunity arising from CRIS and the previously described natural language processing developments (Jackson et al. 2017; Patel et al. 2015) to ascertain symptoms spanning six domains, against which I investigated death and hospitalisation outcomes for a cohort of individuals with SMI testing the null hypothesis that symptom domains are not associated with hospitalisation or mortality .

9.4 Methods

9.4.1 Setting and cohort definition

The setting for this study has been described in chapter 5. Cohort members were defined as cases given a diagnosis of schizophrenia (ICD-10 code: F20.X), schizoaffective disorder (F25.X) or bipolar affective disorder (F31.X) and who were under active review by SLaM services at any time during the period between 1st January 2007 and 31st December 2010. The age of cohort members on 1st January 2011 was calculated from the year and month of birth and individuals aged 15 to 104 years were included. Diagnoses recorded in CRIS were based on the 10th edition of the World Health Organization International Classification of Diseases (ICD-10).

9.4.2 Overview of analysis strategy

As mentioned above, the cohort was defined as individuals who were alive on 1st January 2011. For this cohort, I ascertained the following: symptom profiles prior to that date, covariates at that date, and outcomes over pre-defined periods from that date onwards.

9.4.3 Data sources

Demographic and clinical data- date of birth, gender and ethnicity were extracted in addition to employment, marital status and ethnicity nearest to 1st January 2011. Date of

birth was used to calculate age on 1st January 2011 and managed as a continuous variable. Ethnicity classifications were- “White”, “Black”, “Asian”, “mixed, unknown and others”. Marital status was classified as being “single”, “in a relationship (cohabiting, married or civil partnership)”, “separated”, “not disclosed or unknown”. Employment was classified as “employed”, “unemployed” or “unknown”. The symptoms data obtained comprised the presence (count of occurrences and date first recorded) of any symptom, from the list shown in Table 5.2 (Chapter 5), recorded at any stage from their first contact at SLaM until the end of a pre-defined observation period on 31st December 2010. The specific symptom domains of research interest were classified as positive, negative, disorganisation, manic, catatonic, and depressive, extracted from open-text fields of clinical events and correspondence letters in CRIS. The precision and recall of each symptom are presented in appendix 1. Research papers by Jackson et al 2017 and Patel et al 2015 describe the methodology for identifying symptoms within CRIS. The process, described in more detail in chapter 5, involved using NLP information extraction within clinical records stored in CRIS to process large quantities of unstructured (human authored) text and return structured information on presence of symptoms. As described in chapter 5, a detailed review of the training data set was undertaken to identify and annotate key phrases within the records that were either relevant or irrelevant for keywords related to each symptom with inter-rater reliability between annotators monitored. This training data set was used to construct an application using a hybrid classification model, consisting of a support vector machine learning algorithm and rule-based text matching, using the Generalised Architecture for Text Engineering (GATE) software package.

Mortality data- Data extraction for this data sources is outlined in chapter 5 and described in chapter 6. Deaths occurring in the cohort between 1st January and 31st December 2011 were analysed to measure short term impacts of the mental health conditions.

Hospitalisation (Hospital Episode Statistics, HES) data- Data extraction for this data source is outlined in chapter 5 and described in chapter 7. A hospitalisation was defined as having at least one HES episode recorded. The total number of days in hospital within the observation period between 1st January 2011 and 31st December 2014 was collected to measure short term impacts of the mental health conditions over a four year period and to capture admissions to non-mental health hospitals. The date of admission and date of discharge were used to calculate use of inpatient care services. Hospitalisation data were stratified into whether the admission took place into a mental health or non-mental health Trust.

9.4.4 Statistical analysis

To summarise the measurements for the symptom domains, each specific symptom was coded as "1" for ever being mentioned or "0" for absence of any trace of it in open text fields. The count of mentioned symptoms was then considered as a comprehensive indicator for each domain. This means that an individual who had two different symptoms recorded at any time would score two points, someone who had three different symptoms recorded would score three points, and so on. Standardised Z-scores were calculated for each symptom domain – i.e. the count of symptoms for each domain was divided by its standard deviation in the sample to unify the measurement scale for all domains and facilitate direct comparison.

Cox regression procedures were used to model associations between the exposures of interest (positive, negative, catatonic, disorganisation, manic and depressive symptoms) and all-cause mortality. Proportional hazards assumptions were tested and indicated that it was appropriate to use Cox regression procedures. For each individual, the 'at-risk' period commenced from the date of contact was made with SLaM services during the observation period. The censoring date was the end of the observation period (31st December 2014) and the event date was the date of death during the observation period. Adjustment for age, gender, ethnicity, employment status and marital status as potential confounders to association between mortality and the exposures was carried out.

For the analyses of hospitalisation outcomes, univariate and multivariate linear regressions were applied individually for mental healthcare admissions and for non-mental admissions

with symptom domains as the major exposures of interest. Total length of hospital stay in days and number of admissions between 1st January 2011 and 31st December 2014 were used as dependent variables. Age, gender, ethnicity, employment status and marital status, variables described in chapter 5, were considered as confounders.

9.5 Results

Table 9.1 describes the characteristics of this cohort as at 1st January 2011: 14,824 individuals with a primary diagnosis of SMI who met the inclusion criteria. The age group that included the most individuals was 35-44 years (25.3%) followed closely by 45-54 years (22.4%). Unemployment was noted for 36.8% of the cohort; 65.2% of the cohort were single; 51.6% were from a White ethnic background; 94.5% had at least one positive symptom and 94.0% had depressive symptoms. The lowest prevalence (51.1%) of at least one symptom was noted for catatonic symptoms; 821 individuals (5.5%) had at least one covariate other than date of birth missing. Table 9.2 presents the symptoms extracted within each of the six domains. Among positive symptoms, the most frequent was hallucinations (82.7%) and the least frequent was arousal (39.2%), while for negative symptoms, poor rapport (77.8%) was most common and poverty of thought (6.0%) was least. Among disorganisation symptoms, circumstantiality (77.2%) was most frequent and derailment (9.9%) was least. For manic symptoms, elevated mood (95.9%) was most frequent and insomnia (30.8%) was least. Among catatonic, posturing (27.8%) was most frequent and stupor (1.4%) was least while for depressive symptoms, most and least frequently recorded symptoms were loss of appetite (83.5%) and anergia (3.1%) respectively.

Table 9:1- Cohort characteristics of all individuals with a diagnosis of SMI received prior to 31st December 2010 and had active contact with SLaM between 1st January 2007 and 31st December 2010 (n=14,824)

Variable		Number of individuals
Age on 1st January 2011(years)	15- 24	1017
	25- 34	2867
	35- 44	3747
	45- 54	3317
	55- 64	1867
	65- 74	1133
	75- 84	658
	85- 94	202
	95- 104	16
Gender	Males	7735
	Females	6268
Employment	Unemployed	5455
	Employed	1860
	Unknown/ missing	7509
Marital status	Single	9668
	Married/ civil partnership/ cohabiting	1897
	Divorced/ civil partnership dissolved/ separated	1610

Variable		Number of individuals
	Widowed/ surviving civil partner	506
	Not disclosed/ unknown	1143
Ethnicity	White (British, Irish, any White background)	7654
	Black (African, Caribbean, any Black background)	4941
	Asian (Indian, Chinese, any Asian background)	847
	Mixed/ unknown/ missing	1382
Presence of at least one symptom	Positive	14,015 (94.5%)
	Negative	13, 472 (90.9%)
	Disorganisation	13,323 (89.9%)
	Manic	11,475 (77.4%)
	Catatonic	7,581 (51.1%)
	Depression	13,949 (94.0%)

Table 9:2- Symptom groupings of all individuals with a diagnosis of SMI prior to 31st December 2010 and had active contact at SLaM between 1st January 2007 and 31st December 2010 (n=14,824)

Number of cases with individual symptoms ever recorded prior to 1st January 2011					
Positive symptoms	Negative symptoms	Disorganisation symptoms	Manic symptoms	Catatonic symptoms	Depressive symptoms
Agitation 10195	Blunted or flat affect 11257	Circumstantial speech 11438	Elevated mood 14217	Mannerism 1308	Anergia 458
Aggression 10332	Emotional withdrawal 10608	Reduced coherence 9604	Disturbed sleep 11956	Rigidity 2900	Anhedonia 3404
Arousal 5810	Social withdrawal 9550	Formal thought disorder 10838	Insomnia 4570	Posturing 4121	Loss of appetite 12375
Hostility 6959	Abstract thinking- 2933	Thought block 2182	Pressured speech 13207	Perseverance 2649	Poor concentration 10976
Delusions 11895	Poor rapport 11536	Tangential speech 3951	Irritability 10094	Stupor 212	Low energy 9219
Hallucinations 12267	Apathy 1749	Derailment 1473	Elation 13889	Waxy flexibility 118	Guilt 6292
Paranoia 11920	Poverty of speech 1505	Flight of ideas 3985	Grandiosity 5712	Immobility 559	Helplessness 2430
Persecution 8274	Poverty of thought 893			Echolalia 327	Hopelessness 4908
				Mutism 2546	Poor motivation 10412

Catalepsy 13

Echopraxia 53

Psychomotor retardation 6109

Suicidal ideation 5256

Tearfulness 7459

Weight loss 7264

Worthlessness 2579

Table 9.3 summarises outcome associations of factors that were adjusted for in the subsequent analyses: age, gender, ethnicity, employment status and marital status. The results shown in this table are univariate / unadjusted analysis outcomes. Increasing age was associated with worse outcomes on all three measures- mortality ($p<0.001$), number of admissions ($p<0.001$) and total inpatient care ($p<0.001$). Male gender was associated with higher number of admissions. Employed status was associated with higher mortality (HR 2.40, 95% CI 1.99- 2.89) and significant overall inpatient hospitalisation (B value 0.75, 95% CI 0.05- 1.44). Mortality was highest for individuals known to be widowed (HR 6.48, 95% CI 5.27- 7.96). Widowed individuals also experienced higher numbers of admissions (B value 0.59, 95% CI 0.20- 0.97). individuals from all other ethnic groups compared with White ethnic background experienced less mortality. Asians also experienced lower number of admissions while individuals from a Black ethnic group experienced shorter stays as inpatients.

Table 9:3- Univariate analyses on the outcomes of mortality, number of non-mental health related admissions and total inpatient care at non-mental health hospitals

Variable	Mortality HR (95%), p-value	Admission number B value (95%), p-value	Total inpatient care (days) B value (95%), p-value
Age (per year increase)	1.08 (1.07- 1.09), p<0.001	0.02 (0.01- 0.03), p<0.001	0.06 (0.04- 0.72), p<0.001
Gender			
Males	REF	REF	REF
Females	0.90 (0.78- 1.02), p=0.11	-0.19 (-0.32- -0.06), p<0.001	-0.02 (-0.45- 0.41), p=0.93
Employment status			
Unemployed	REF	REF	REF
Employed	2.40 (1.99- 2.89), p<0.001	0.17 (-0.04- 0.38), p=0.11	0.75 (0.05- 1.44), 0.04
Unknown/ missing	1.14 (0.97- 1.33), p=0.10	0.06 (-0.09- 0.20), p=0.43	0.48 (0.00- 0.95), 0.05
Marital status			
Single	REF	REF	REF
Married/ civil partnership/ cohabiting	1.32 (1.08- 1.61), p=0.01	-0.17 (-0.37- 0.03), 0.10	-0.41 (-1.06- 0.25), p=0.23
Divorced/ civil partnership dissolved/ separated	1.59 (1.30- 1.95), p<0.001	0.13 (-0.08- 0.35), p=0.22	0.11 (-0.59- 0.81), p=0.77
Widowed/ surviving civil partner	6.48 (5.27- 7.96), p<0.001	0.59 (0.20- 0.97), p<0.001	1.01 (-0.26- 2.27), p=0.12

Variable	Mortality HR (95%), p-value	Admission number B value (95%), p-value	Total inpatient care (days) B value (95%), p-value
Not disclosed/ unknown	0.91 (0.68- 1.22), p=0.53	0.06 (-0.19- 0.32), p=0.62	-0.87 (-1.71- -0.02), p=0.04
Ethnicity			
White (British, Irish, any White background)	REF	REF	REF
Black (African, Caribbean, any Black background)	0.50 (0.43- 0.59), p<0.001	-0.10 (-0.25- 0.04), p=0.17	-0.47 (-0.95- 0.01), p=0.05
Asian (Indian, Chinese, any Asian background)	0.61 (0.44- 0.84), p<0.001	-0.37 (-0.64- -0.09), p=0.01	-0.68 (-1.59- 0.24), p=0.15
Mixed/ unknown/ missing	0.40 (0.30- 0.55), p<0.001	-0.24 (-0.47- -0.01), p=0.04	-0.40 (-1.16- 0.36), p=0.30

Table 9.4 describes the findings from Cox regression analyses of domain-specific symptom scores in relation to mortality during 2011. Examining proportional hazards assumptions for each symptom domain as an exposure, no significant deviations from this assumption were observed (p-values ranging from 0.7 to 0.9). After adjusting for age, sex, employment, marital status and ethnicity, higher levels of positive symptoms, negative symptoms and catatonic symptoms were associated significantly with increased risk of mortality with similar adjusted hazard ratios (1.08 (95% CI 1.03- 1.16), 1.09 (95% CI 1.02- 1.16) and 1.09 (95% CI 1.03- 1.16) respectively). For disorganisation, manic and depressive symptom domains, the hazard ratios were weaker (1.04 (95% CI 0.98- 1.11), 1.02 (95% CI 0.96- 1.08) and 1.02 (95% CI 0.96- 1.09) respectively) and not statistically significant. The correlation between symptom domains is presented in appendix 5 table 1.

Table 9:4- Cox regression analysis of mortality between 1st January 2011 to 31st December 2011 in individuals with SMI according to domain-specific symptoms recorded prior to 2011 - positive, negative, disorganisation, manic, catatonic and depressive (n=865 deaths)

Symptom domain score	Crude Hazard Ratio* (95% CI)	p-value	Adjusted Hazard Ratio* (95% CI)	p-value
Positive	0.93 (0.87- 0.99)	0.03	1.08 (1.03- 1.16)	0.03
Negative	0.92 (0.87- 0.97)	<0.01	1.09 (1.02- 1.16)	0.04
Disorganisation	0.87 (0.82- 0.92)	<0.01	1.04 (0.98- 1.11)	0.16
Manic	0.91 (0.86- 0.96)	<0.01	1.02 (0.96- 1.08)	0.51
Catatonic	1.00 (0.95- 1.05)	0.95	1.09 (1.03- 1.16)	0.01
Depressive	0.87 (0.82- 0.92)	<0.01	1.02 (0.96- 1.09)	0.41

*Hazard ratios were derived from exposures entered as z-transformed independent variables – i.e. entered in standard deviation units. Adjustments were for age, gender, ethnicity, employment status and marital status

Associations between symptom levels and mental health hospitalisations are displayed in Table 9.5. with the primary aim of establishing construct validity before moving on to analyses of non-mental health hospitalisation. Mental health admissions were significantly raised in association with higher scores on each of the symptom domains analysed; combined mental healthcare inpatient days were also significantly increased. After adjustment, the regression model for combined inpatient days showed the largest effect size with catatonic symptoms (B-value 9.95, 95% CI 9.04- 10.9) and the smallest with manic symptoms (B-value 3.12, 95% CI 2.18- 4.06). After adjustment, associations with higher numbers of admissions were significant for positive (B-value 0.22, 95% CI 0.20- 0.24) , disorganisation (B-value 0.22, 95% CI 0.20- 0.24) and catatonic (B-value 0.22, 95% CI 0.20- 0.23) symptoms previously recorded.

Table 9:5- Linear regression models of number of mental healthcare inpatient admissions and total inpatient days in patients with SMI during 2011-2014 by levels of symptoms recorded prior to 2011- positive, negative, disorganisation, manic, catatonic and depressive (n=14,824)

	Number of admissions				Total inpatient care (days)			
Symptom domain score	Crude* B-value (95% CI)	p-value	Adjusted* B-value (95% CI)	p-value	Crude* B-value (95% CI)	p-value	Adjusted* B-value (95% CI)	p-value
Positive	0.24 (0.22- 0.25)	< 0.01	0.22 (0.20- 0.24)	< 0.01	6.73 (5.84- 7.62)	< 0.01	6.61 (5.64- 7.57)	< 0.01
Negative	0.16 (0.14- 0.17)	< 0.01	0.12 (0.11- 0.14)	< 0.01	4.84 (3.94- 5.73)	< 0.01	4.39 (3.44- 5.35)	< 0.01
Disorganisation	0.25 (0.23- 0.26)	< 0.01	0.22 (0.20- 0.24)	< 0.01	6.43 (5.54- 7.32)	< 0.01	6.12 (5.17- 7.07)	< 0.01
Manic	0.15 (0.13- 0.16)	< 0.01	0.12 (0.11- 0.14)	< 0.01	3.39 (2.49- 4.28)	< 0.01	3.12 (2.18- 4.06)	< 0.01
Catatonic	0.24 (0.22- 0.26)	< 0.01	0.22 (0.20- 0.23)	< 0.01	10.0 (9.15- 10.9)	< 0.01	9.95 (9.04- 10.9)	< 0.01
Depressive	0.16 (0.15- 0.18)	< 0.01	0.14 (0.12- 0.16)	< 0.01	3.62 (2.73- 4.51)	< 0.01	3.66 (2.71- 4.61)	< 0.01

* Hazard ratios were derived from exposures entered as z-transformed independent variables – i.e. entered in standard deviation units. Adjustments were for age, gender, ethnicity, employment status and marital status

In analyses of admissions to non-mental health hospitals (Table 9.6), higher manic (B value= 0.09, 95% CI 0.02- 0.15), catatonic (B value= 0.08, 95% CI 0.02- 0.15) and depressive (B value=0.14, 95% CI 0.08- 0.21) symptoms were significantly associated with higher numbers of admissions after adjustment. No significant associations were found for positive, negative and disorganisation symptoms (B-values 0.05, 0.06 and 0.06 respectively). No associations were found in adjusted models between any of the symptom domains and combined inpatient days.

Table 9:6- Linear regression analyses of general hospital admissions (all hospitals not identified as a mental health trust) and total inpatient care during 2011-2014 in patients with SMI by levels of symptoms recorded prior to 2011- positive, negative, disorganisation, manic, catatonic and depressive (n= 14,824)

	Number of admissions				Total inpatient care (days)			
Symptom domain score	Crude B-value (95% CI)	p-value	Adjusted* B-value (95% CI)	p-value	Crude B-value (95% CI)	p-value	Adjusted* B-value (95% CI)	p-value
Positive	-0.01 (-0.07- 0.05)	0.72	0.05 (-0.01- 0.12)	0.12	-0.09 (-0.30- 0.11)	0.38	0.06 (-0.16- 0.29)	0.58
Negative	-0.02 (-0.08- 0.04)	0.55	0.06 (-0.01- 0.13)	0.09	-0.01 (-0.22- 0.20)	0.94	0.18 (-0.04- 0.41)	0.10
Disorganisation	-0.02 (-0.09- 0.04)	0.45	0.06 (-0.01- 0.12)	0.10	-0.21 (-0.42- -0.01)	0.04	-0.03(-0.25-0.19)	0.80
Manic	0.03 (-0.03- 0.09)	0.35	0.09 (0.02- 0.15)	0.01	-0.09 (-0.30- 0.11)	0.38	0.02 (-0.19- 0.24)	0.84
Catatonic	0.03 (-0.03- 0.09)	0.29	0.08 (0.02- 0.15)	0.01	-0.04 (-0.25- 0.17)	0.72	0.08 (-0.14- 0.29)	0.48
Depressive	0.06 (0.00- 0.13)	0.05	0.14 (0.08- 0.21)	<0.01	-0.10 (-0.31- 0.11)	0.34	0.06 (-0.16- 0.28)	0.58

*adjusted for age, gender, ethnicity, employment status and marital status

9.6 Discussion

Using a large mental healthcare EHR data source and extensive NLP-derived profiling of symptomatology in SMI, novel information on fifty-three symptoms across six symptom domains was extracted and analysed against mortality and hospitalisation as outcomes. Whilst primarily this was an analysis of physical health outcomes focusing on mortality and general hospital use, the associations with mental healthcare hospitalisation outcomes (both number of episodes and total duration) supported the construct validity of the symptoms. After adjustment, higher scores on positive, negative and catatonic symptom domains were significantly associated with an increased risk of mortality over a 12-month follow-up period, but hazard ratios were weaker and not significant for disorganisation, manic and depressive symptom domain scores. For admission to non-mental health hospitals, only manic, catatonic and depressive symptoms were associated with higher numbers of admissions and no symptom domain score predicted combined duration. In the early study on negative symptoms described in chapter 8, increased likelihood of mental health hospital admissions, duration of admissions and increased likelihood of readmission following discharge were associated with negative symptoms for a cohort of individuals with schizophrenia. Further developments have now enabled the investigation of a more comprehensive suite of symptoms to investigate physical health outcomes across SMI more broadly.

In relation to the literature reviewed in chapter 2, the extraction of this novel information allowed a more comprehensive investigation of the association between SMI symptom

profiles and physical health outcomes than has previously been attempted. Considering adjusted models for mortality, symptoms most strongly associated with this outcome were the more schizophreniform domains, i.e. positive, negative and catatonic. Further investigations may identify whether people with more affective symptoms engage better with healthcare or receive more targeted care that reduce adverse outcomes. For example, a study looking at myocardial infarction in patients with SMI found that both schizophrenia and bipolar disorder were associated with less receipt of post-event care but only schizophrenia was associated with higher short-term mortality (Wu S-I, et al 2013). Studies of SMI cohorts to date have not investigated as wide a range of symptoms as described in this study, although there are some links to mortality already identified. In a prospective cohort study among individuals with bipolar disorder, clinically significant manic/hypomanic symptoms predicted cardiovascular mortality independent of diagnosis, treatment exposure, age, gender, and cardiovascular risk factors at intake; however, depressive symptom burden was not related to cardiovascular mortality, having measured the severity of affective psychopathology using weekly Psychiatric Status Ratings of symptom levels (Fiedorowicz et al. 2009). That study considered proportions of weeks exposed during follow-up for symptom variables, in order to reduce limitations from loss to follow-up; this was not a particular issue for my study because follow-up will have been near-complete with the linkages to national data sources. The number of individuals followed was also orders of magnitude higher in my cohort: the study reported by Fiedorowicz followed 435 individuals recruited from a research cohort, whereas my cohort comprised more than 14,000 individuals with SMI, albeit with more limited information available. In another study limited to assessing outcomes for manic symptoms, after adjusting for major depressive symptoms and demographic differences, the odds of mortality at follow-up for participants

with lifetime manic spectrum episodes in the 30-44 and 45-64 year age cohorts at baseline were higher than those with no lifetime manic spectrum episodes in the same age cohorts, with odds ratios of 1.39 and 1.41 respectively (Ramsey et al. 2013).

Pre-structured information in UK mental health records on a limited range of symptoms is available from HoNOS: a set of scales to measure the health and social functioning of people with severe mental illness (Wing JK, Beevor AS, Curtis RH 1998). HoNOS is the most widely used routine clinical outcome measure used by English mental health services but the insight it provides as an exposure is limited to specific items measuring behaviour, impairment, symptoms and social functioning. HoNOS data were previously investigated as exposures in analyses of all-cause mortality using CRIS data reported by Hayes RD et al 2011. The key findings from that study were that mortality was not significantly associated with the HoNOS "hallucinations and delusions" or "overactive-aggressive behaviour" scales which in my study would be included under the positive symptoms domain. There was a significant positive association between mortality and subclinical depression as recorded on HoNOS among individuals with schizophrenia (adjusted hazards ratio of 1.5) but a negative association with subclinical and more severe depression among patients with schizoaffective disorder (adjusted hazard ratios 0.10 and 0.30 respectively) (Hayes et al. 2012). However, HoNOS provides very limited information on symptom profiles and certainly does not cover all the domains discussed here. In my analysis, no association was found between the depressive symptoms measure and mortality.

For non-mental health hospitalisation, my findings suggest that any effect is on the number of admissions rather than their duration and an opposite pattern from mortality was observed in that the manic and depressive symptom scores were the strongest predictors. These exposures might therefore be increasing the risk of multiple, short duration admissions rather than long-duration stays which would fit with the association with number of admissions but not total duration. People distressed by their SMI often do not seek out services or choose to fully engage in them (Corrigan et al. 2014), and one factor that impedes care seeking and undermines the service system is mental illness stigma. However, it is unclear whether medical help-seeking behaviour varies within SMI – for example, whether individuals who experience manic symptoms to have a greater propensity for medical help-seeking. Catatonic symptoms were associated with a higher risk of both mortality and number of admissions, although it should be borne in mind that not all catatonic symptoms mentioned in text will have been specific to 'catatonic syndrome', and at least some might reflect other phenomena such as adverse drug events and/or might have occurred secondary to underlying physical health conditions. Therefore, further investigation of this issue is required before definitive inferences can be made.

Associations between symptom profiles and general hospitalisations are under-researched in that they are either overlooked or only ascertained in conjunction with mental health admissions. For example, a study investigating whether better primary care can reduce hospital admissions used the SMI quality indicators measured routinely in the Quality and Outcomes Framework (QOF) in English general practice (Jacobs et al. 2015). The study tested the impact of SMI QOF indicators on length of stay and its associated costs using

linked administrative data from around 8,500 general practitioner (GP) practices and from Hospital Episode Statistics. SMI admissions were defined by a main ICD-10 diagnosis of F20–F31, and covariates included information on GP practice and patient population characteristics, area deprivation and other potential confounders such as access to care. Analyses carried out at a GP practice level for admissions found a positive association between QOF achievement and admissions: i.e. found that practices with better QOF markers also had higher admissions. This was observed for emergency admissions both for mental and physical health, with an additional 10% in QOF achievement associated with an approximately 1.9% increase in the practice emergency SMI admission rate (Jacobs et al. 2015). However, there was no significant association of QOF achievement with either length of stay or cost. The authors suggested that higher quality primary care, as measured by QOF may not effectively prevent the need for secondary care but acknowledged that they were not able to confirm whether patients received their QOF checks prior to admission rather than post-discharge. Better quality primary care may also be picking up unmet need for secondary care, and QOF measures may not accurately reflect quality of primary care. In another study evaluating the relationship between baseline levels of and changes in Positive and Negative Syndrome Scale (PANSS) clinical symptom scores and risks for subsequent psychiatric hospitalisation, reductions in PANSS total scores during a three-month period were associated with a significantly lower risk for psychiatric hospitalisation with an odds ratio of 0.78 (Glick et al. 2015). Ten-point reductions in PANSS total score during three months were also associated with reduced predicted number of psychiatric hospitalisations and inpatient care (number of nights in hospital) over a three-year period.

I believe that my study provides unique contributions to the evidence base, being one of the first to investigate detailed SMI symptom profiles as predictors of physical health outcome measures. Key strengths of the study included the large cohort size and comprehensiveness of the patient sample, and its likely representativeness of people with SMI in its local catchment, since SLaM is a near-monopoly provider of secondary mental health care. Electronic health records are available for all patients in SLaM and therefore, it was possible to include many records for symptoms extraction. This provided high levels of statistical power whilst also enabling adjustment for key demographic factors. Using nationally available data sources such as Office for National Statistics mortality data as well as Hospital Episode Statistics meant that mortality data and hospital admission data were comprehensive and reduced the potential for bias arising from incomplete follow-up. Coverage is anticipated to be good for both these linked data sources as there is a legal requirement to record every death, while the hospitalisation data have an administrative function and are linked to payments to hospitals. The linkage to a national data source also meant that loss to follow up is very limited unlike that described in other studies (Fiedorowicz et al. 2009).

There are limitations to this study. From a study design perspective, there are two main issues to consider. Firstly, even with the approach taken of ascertaining symptoms recorded at any stage, detection bias is likely; for example, individuals who died soon after being seen by the mental health service will have fewer opportunities to have symptoms recorded on their record, and there was no attempt to account for the duration of patient records. Secondly, there is likely to be underestimation of some symptoms. For example, it is

possible to have missed some aspects of a psychosis presentation – either because particular symptoms were systematically missed, or because of target symptoms included in this study were described in non-standard language (e.g. 'hearing voices', rather than 'auditory hallucination'). Similar to Fiedorowicz (2009) measures of symptoms in my study can only approximate cumulative lifetime exposure as I used extracted symptoms recorded at any stage in the patient record. It is also possible that the SMI diagnosis had been first recorded at an earlier presentation and at least some members of the cohort might have a different presenting complaint (e.g. an individual with a history of schizophrenia with reported symptoms now currently receiving care for alcohol or drug dependence). The analysis presented in this chapter did not differentiate affective vs. non-affective psychosis but instead aimed to characterise symptom profiles across SMI in order to investigate outcomes. Considering symptoms, within the catatonia domain, further exploration of the domain is required before making inference for the reasons mentioned previously. Mortality was measured over quite a short time-period and therefore, the outcome might reflect shorter term impacts of mental health conditions. Further, the study was limited to assessing the variation in access to secondary care because Hospital Episode Statistics were used as the data source, and did not incorporate any information from primary care. HES as a data source also may not reflect all healthcare utilisation as some individuals may access private health care, although this is likely to have been rare for the SLaM catchment. In addition, the hospitalisation component of this study could only look at met need. This study did not attempt to split hospitalisation outcomes into emergency and elective care. However, people with SMI use more emergency care than those without: for example, a review of admission patterns found that individuals with SMI attended the emergency department 3.2 times more often and experienced 4.9 times as many emergency inpatient

admissions than those without mental ill health in 2013/14. In addition, for elective care, people with mental illness used less planned inpatient care than those without. (Dorning, H. Davies, A. Blunt 2015).

Despite these limitations, this study comprising data from a large number of individuals with SMI provides meaningful implications. Symptoms data might, for example, be used to profile behavioural traits that could impact on preventative measures such as screening programmes, or might be investigated as factors influencing the effectiveness of interventions. There are also more direct financial implications as measures to reduce severity of symptoms or minimise the development of new symptoms may mean less costs of hospitalisation. No economic analyses were attempted to quantify this; however, in a US study, a 10-point reduction in Positive and Negative Syndrome Scale (PANSS) total score was associated with a savings in psychiatric hospitalisation cost of $192 over three months and $1,135 over a year (Glick et al. 2015). Reductions in PANSS total scores significantly reduced risks for psychiatric hospitalisations, total number of psychiatric hospitalisations, total nights for psychiatric admissions, and the costs of these hospitalisations highlighting the benefits of symptom control on the direct costs of care in schizophrenia (Glick et al. 2015). Policy implications include a need to recognise and address the fact that, while collectively individuals with SMI can be grouped together by similarities, specific pathophysiological links between sub-groups of syndromes can impact some individuals and not others. From a research perspective, this study provides a precedent for a new wave of studies utilising in-depth information on large routine clinical databases to understand better the key issues pertinent to the health care of individuals with SMI and other mental disorders at least. In

addition to further refining the NLP applications developed to date to maximise their potential, there may also be opportunities to explore the measurement of functioning in a more thorough manner through applying similar NLP techniques, improving on than what is currently available through scales such as HoNOS.

9.7 Summary of findings from chapter

NLP-extracted symptoms of SMI showed construct validity in their robust prediction of adverse mental health outcomes. Considering mortality and non-psychiatric hospitalisation, higher levels of schizophreniform symptoms (positive and negative) were associated with increased risk of mortality, while higher levels of manic and depressive symptoms were associated with higher numbers of hospitalisations. Higher levels of catatonic symptoms were associated with both of these outcomes. Total duration of hospitalisation within the observation period was not predicted by any symptom scale.

10 Discussion

10.1 Chapter outline

Having considered all findings from previous chapters in relation to existing research, this chapter aims to summarise the key findings in the context of the broader strengths and limitations of the thesis and consider how my findings fit into the wider context of care for individuals with serious mental illnesses. I discuss other possible implications and ideas for further research.

10.2 Summary of findings

Overall, this thesis provides evidence that using linkages and electronic health records, it is possible to understand the physical health issues that are pertinent to the morbidity and mortality of individuals with SMI. The objectives of this thesis are described and discussed below.

Objective 1- To describe the relative contributions of major disease groups to the gap in life expectancy between individuals with SMI and the general population.

This analysis aimed to estimate contributions of specific causes of death to the gap between people with SMI and the general population. Natural causes accounted for 79.2% of lost life-years in women with SMI and 78.6% in men. Deaths from circulatory disorders accounted for more life-years lost in women than men (22.0% versus 17.4%, respectively), as did deaths from cancer (8.1% versus 0%), but the contribution from respiratory disorders was lower in women than men (13.7% versus 16.5%). For women, cancer contributed more in those with non-affective than affective disorders, while suicide, respiratory and digestive disorders contributed more in those with affective disorders. In men, respiratory disorders contributed more in non-affective disorders. Other contributions were similar between gender and affective/non-affective groups. Loss of life expectancy in people with SMI was thus accounted for by a broad range of causes of death, varying by gender and diagnosis.

Objective 2- To describe the most common reasons for admission to non-psychiatric hospitals by individuals with SMI and what are the relative frequencies of these admissions compared to the general population.

This analysis aimed to delineate frequencies of and causes for non-psychiatric hospital admissions in SMI and compare to general population in the same area. The commonest discharge diagnosis categories in the SMI cohort were urinary conditions, digestive conditions, unclassified symptoms, neoplasms, and respiratory conditions. Standardised admission ratios were raised for most major categories, except neoplasms for which there was a significantly lower risk in SMI compared to the general population. Hospitalisation risks were specifically higher for poisoning and external causes, injury, endocrine/metabolic conditions, haematological, neurological, dermatological, infectious and non-specific ('Z-code') causes. The five commonest specific ICD-10 diagnoses at discharge were 'chronic renal failure' (N18), a non-specific code (Z04), 'dental caries' (K02), 'other disorders of the urinary system' (N39) and 'pain in throat and chest' (R07), all of which were higher than expected (standardised admission ratios ranging 1.57-6.66). A range of reasons for non-psychiatric hospitalisation in SMI were thus apparent, with self-harm, self-neglect and/or reduced healthcare access, and medically unexplained symptoms as potential underlying explanations.

Objective 3- As a proof of principle for ascertaining meaningful symptom profiles from routine mental health record text fields, to describe the prospective association between the number of recorded negative symptoms and mental healthcare outcomes (admission, duration of admission, and readmission) among individuals with schizophrenia.

This analysis aimed to use NLP to develop a suite of language models to capture key symptoms of SMI from clinical text, to facilitate the secondary use of mental healthcare data in research. Using a Support Vector Machine learning method with an NLP tool, it was possible to successfully extract data on negative symptoms from the electronic mental health records of a large clinical sample of patients with schizophrenia. This approach enabled the use of data recorded as part of routine clinical care for analyses to investigate associations with adverse health outcomes. Negative symptoms of schizophrenia ascertained in this way were scaled, evaluated, and found to be associated with increased risk of psychiatric hospitalisation, longer duration of inpatient episodes, and higher risk of readmission following discharge.

Objective 4- To describe the association with mortality and hospitalisation for each of six symptom dimensions (positive, negative, manic, disorganisation, catatonic and depressive) extracted from the clinical records of individuals with SMI.

Having ascertained symptom mentions in the electronic clinical records of a large cohort of individuals with SMI, this analysis was carried out to assess its associations with subsequent all-cause mortality and hospitalisation. In a cohort of over 14,000 SMI cases, higher numbers of positive, negative and catatonic symptoms were associated with higher mortality. Higher numbers of non-mental healthcare hospital admissions were significantly associated with manic, catatonic and depressive symptom domains, but no associations were found with total inpatient care after adjustment.

10.3 Strengths and limitations

I have considered specific strengths and limitations in the chapters covering individual component analyses but will now consider some broader strengths and limitations which need to be borne in mind when interpreting thesis findings.

The major strength of this PhD study is the novel methodology using electronic secondary mental health care data with linkages to mortality and hospitalisation data. CRIS is a unique resource which allows access to “real world” clinical data. Large studies have typically been carried out using information from registries and comparatively CRIS provides beyond this. The more recent advances to information ascertainment through natural language processing techniques meant that I could obtain information on an extensive list of symptoms which has not been included in previous studies on this topic. It also meant that I had larger sample sizes than conventional studies and more informative clinical notes as a data source compared with administrative data sources.

An important limitation to note is that the primary purpose of an EHR is to record patient care; thus, rich and diverse contextual information is potentially available but it has not been recorded for research purposes and there may be inaccuracies, and factors influencing recording, which need to be borne in mind; in addition, clearly much of the information resource is embedded in free-text fields requiring extraction. Throughout this thesis, recognising that it is not collected under research standards, higher quality information was

sought by focusing on fields with higher clinical relevance and/or where there are higher incentives for accuracy. This, for example, was a key justification for developing multiple NLP algorithms for symptom profiling, as symptoms tend to be recorded for purely clinical reasons, whereas diagnostic codes represent fields which are compulsory and driven to some extent by management priorities, so may suffer in accuracy (particularly when it comes to updating diagnostic changes) as a consequence.

Considering bias, confounding and chance- systematic error could have been introduced as a result of design or conduct of the study leading to an inaccurate estimate of the association between exposure and outcome. The data were obtained from historical cohorts as opposed to a retrospective study. However, the inclusion criteria for the individual analyses may have introduced selection bias. Some loss to follow up may have taken place but the impact is reduced by including two data sources with national coverage. ONS captured anyone who died in the UK and HES identified anyone who was admitted to any hospital across the whole of England. Loss due to migration from the UK is a possibility and must be noted for this study as the catchment area has relatively higher proportion of minority ethnic groups who were born outside the UK and who might therefore have been more likely to emigrate.

Misclassification bias occurs when categorisation is not accurate with respect to exposure or outcome. Random misclassification is likely to underestimate an effect. However, a non-random misclassification, where an exposure is dependent on the outcome or outcome is

dependent on the exposure, could over- or underestimate effects. More unwell individuals have more clinical documentation, where symptoms are more likely to be recorded. In essence, this describes a similar concept to that of confounding by indication where symptoms appear to be most comprehensively recorded in the most unwell patients. This study could not assess disease severity to able to differentiate presentations between more or less unwell individuals. However, the large sample size was at least able to include individuals receiving mental health services regardless of whether this involved inpatient or outpatient care, unlike the large number of hospitalisation-based registries cited in Chapter 2. Also, there is less measurement bias at the point of clinical care as clinical staff entering information during clinical care are not aware of the research use of the data. At the stage of symptom applications development where training sets were developed and the clinical team were annotating individual symptoms, there is possibility of introducing a misclassification bias. This was handled with at least two team members conducting independent annotations and inter-rater agreement measured for each item. Also, while the annotator was aware the case notes were for an individual with SMI, they did not know the exact diagnosis or any other information other than the sentences on either side of the annotating word or phrase.

Adjustments were made for key demographic factors and the large sample sizes allowed multivariable analyses. However, there were several variables that were not included in the analyses which I think would be useful to evaluate further, including individual lifestyle factors such as smoking, medication use as well as area-level factors such as deprivation. Also, the non-randomised design inevitably does not allow residual confounding to be

excluded. Prospective recording of data on the EHR meant that reverse causality is less likely.

The large sample sizes assisted in considering the role of chance by providing small standard errors around estimates. However, it should be borne in mind that relatively small effects could therefore be identified as statistically significant, and clinical importance should not be assumed for all significant findings. Considering type 2 statistical error (i.e. the false rejection of the null hypothesis), multiple comparisons increase the risk of this and should be viewed with caution. For this reason, numbers of outcomes were minimised to those felt to be most clinically important (e.g. mortality, hospitalisation risk etc.) and exposures such as symptoms were grouped into dimensions rather than analysed individually, except for explicitly exploratory or confirmatory investigation.

The generalisability of analyses must be noted as the findings clearly refer to individuals with SMI who had made contact with secondary mental health services within the specified time period. However, as the study is based in a secondary healthcare setting with near-monopoly provision to a geographic catchment, it is unlikely to bias findings to a large extent as SMIs by nature are disorders where most cases will have received secondary care input (NICE 2014). SLaM itself serves an inner and outer urban catchment, so there may be limited generalisability to rural areas; furthermore, its academic status and some specialist service provision might limit generalisability to more generic services.

10.4 Implications

10.4.1 Research implications

This study built on innovative methodology using a comprehensive secondary mental health data source. As described in Chapter 2, large databases used for investigating, in this instance, mortality outcomes in SMI have tended to be limited by the data extracted from administrative resources (or else have involved original data collection but on much smaller samples). The primary purpose of the developments described was to improve the depth of information available on individuals with SMI from the EHR, as administrative information resources (including CRIS at its initial development in 2008) tend to contain little information beyond a diagnosis and basic demographic details. Until now, within mental health EHRs, symptom profiles have therefore been 'invisible' when it comes to deriving data for research, service development or clinical audit. Given that symptoms are key determinants of interventions received and outcomes experienced, this has been a major deficiency. Therefore, the methodology developed in this study at least opens up tools/techniques to use the large amounts of SMI symptomatology data contained within electronic health record systems, and provide new insight into the value of using SMI symptoms as predictors of a range of outcome measures. Further research is clearly required to capitalise on this resource – in particular, to clarify the best ways to group the >70 individual symptoms currently identifiable and their distributions (e.g. number/frequency of annotation points) in individual records. Another line of research and development might also include further improvements to define time periods for exposure

measures. This would enable the more effective capture of underlying constructs, as would the ability to estimate symptom severity and/or trajectory.

The HES data linkage allowed a novel investigation of non-mental health hospitalisations for people with mental disorders, supplementing the extensive work on mortality outcomes reviewed in Chapter 2, and surprisingly representing one of the first investigations of its kind. Assuming that other linkages with mental healthcare datasets can be set up, the degree of replicability across sites and services would be an important next step for future research. For example, the specific association with hospitalisations related to sickle cell disease is likely to be related to the ethnic group profile in SLAM's catchment and other catchments might have their own unique features. Of relevance, the CRIS system is now being deployed at three additional Trusts through the D-CRIS programme, and to a number of others through the UK-CRIS initiative, to significantly extend this research capability. Although HES linkages have not yet been set up at other sites, the work described here might in time be more widely re-evaluated in other mental health services, improving the precision of estimates and clarifying levels of between-site heterogeneity.

Future studies could be usefully designed to include more extensive cohorts with adjustment for additional confounders. Higher capture rates of lifestyle-related factors (e.g. smoking, physical activity, diet etc.) might be achievable through novel data linkages or targeted supplementary data collection in targeted groups, as well as through further NLP development. This would help in the assessment the other key confounding or mediating

factors that might determine physical morbidity and mortality in SMI beyond symptom profiles – factors highlighted as deficiencies of previous research in Chapter 2 but which this thesis remained unable to address.

Further to this, the developments to date on symptom ascertainment could also be more widely used within CRIS to address other study questions, addressing the broader question of whether symptom profiles and derived 'phenotypes' provide better predictive data than diagnostic categories. As discussed in Chapter 1, the relative utility of a dimensional rather than categorical classification of mental disorders has been an area of controversy for as long as psychiatric research and clinical practice have been in existence. On the one hand, genetic and imaging evidence has tended to support cross-diagnostic dimensions as more realistic ways to conceptualise psychopathology; on the other hand, clinical practice requires diagnoses for communicating complex symptom profiles. Ultimately the two approaches can only really be compared with respect to their predictive validity, but this has been challenging to date because of the lack of fine-grain detail across sufficient samples with sufficient representativeness. Symptoms recorded in EHR text fields remain a suboptimal reflection of individual subjective experiences; however, they do at least present a finer-grain measurement than diagnostic codes in administrative data resources, and large sample sizes might provide at least some opportunities for head-to-head comparisons.

As described in Chapter 1, there are important recognised health inequalities experienced by people with SMI, and as described in Chapter 2, these remain relatively under-

investigated in terms of causal pathways – e.g. characteristics of people with SMI who are most at risk, or adverse health states responsible for life expectancy differences. My findings in Chapters 6 and 7 suggest that there should be an increasing focus on identifying and evaluating appropriate models for delivering more effective care, particularly, to reduce the gap in life expectancy and reduce admissions for the conditions for which the standardised admission ratios were high. Although some of this involves the design and evaluation of novel interventions to improve physical health outcomes, there is also likely to be a need to develop further observational studies to help define at risk groups and pathways of causation. For example, findings in Chapter 9 suggest that symptom profiles are much stronger predictions of mental health outcomes compared to mortality or acute care hospitalisation, supporting the earlier findings for negative symptoms specifically, reported in Chapter 8. Therefore, there are still questions to be answered concerning whether people with SMI are simply at a generically high risk of adverse physical health outcomes or whether there are sub-groups at particularly high risk who might be more effectively targeted by future preventative interventions.

10.4.2 Clinical and policy implications

My literature review described in Chapter 2 concluded that there is consistent evidence that all-cause mortality for individuals with schizophrenia and bipolar disorder is higher than that observed in the general population. My study reported in Chapter 6 found that the gap in life expectancy was accounted for by a wide range of recorded causes of death, rather than

just one or two predominating ones, with moderate variation in the pattern of contribution by gender, and between people with affective and non-affective SMI diagnoses. When assessing the specific contribution from each grouped cause of death to life expectancy, it was difficult to identify a single group or two. On a similar vein, a range of reasons for non-psychiatric hospitalisation was apparent in findings reported in Chapter 7, with self-harm, self-neglect and/or reduced healthcare access, and medically unexplained symptoms as potential underlying explanations. These findings on hospitalisation rates and causes of death support awareness-raising initiatives of physical conditions leading to early death and hospitalisation. Because this represents multiple physical-mental health interfaces, such awareness-raising needs to be multifaceted, directed to mental healthcare, physical healthcare and primary care, and to the range of professionals working in these sectors, as well as to policy makers. This may include packages of care to help improve lifestyle, diet and proper medication management.

As described in this thesis, EHRs and routine data sources have enabled research utilising rich and diverse contextual information but have drawbacks in that the primary purpose of an EHR is to record patient care and not recorded for research purposes with inaccuracies, and factors influencing recordings. Considering future improvements in data feeding these platforms, its quality is clearly only as good as the information entered in the first place on the EHR. Imposed initiatives to improve data quality are likely to have limited traction with service staff; however, highlighting output from clinical records directly to clinical teams might in itself be educational and result in better quality of recording. This may be best focused at a team or individual level, for example through clinical dashboards, or

enhancements in the summary displays on EHR systems. As an example, any demonstration that symptom profiles have prognostic significance beyond and across diagnostic categories might justify the automated profiling of individuals real-time from their records, which might in turn promote the more accurate recording of symptoms through a virtuous circle. This is particularly relevant as symptom dimensions rated at presentation have shown to be better at predicting outcomes such as time to first remission than a categorical diagnosis (Ajnakina et al. 2018). Furthermore, diagnostic categories in Psychiatry have changed over the years based on both clinical and biological evidence. If a cross-diagnostic construct were to prove to have higher clinical utility than current diagnostic categories then there would be a reasonable argument to re-appraise traditional nomenclature.

In Chapter 7, standardised admission ratios were raised for SMI with respect to most major disease categories (except neoplasms) with specifically higher hospitalisation risks for poisoning and external causes, injury, endocrine and metabolic conditions, haematological, neurological, dermatological, infectious and non- specific ('Z-code') causes. The five commonest specific ICD-10 diagnoses at discharge were 'chronic renal failure', a non-specific code, 'dental caries', 'other disorders of the urinary system', and 'pain in throat and chest', all of which were higher than expected (SARs ranging 1.57–6.66). Focussed policy and interventions to minimise unnecessary non-psychiatric hospitalisation are required. Trials to identify appropriate interventions could be a next step, although it is likely that a complex and supportive strategy to assess and address contextual issues and limitations in patient capacity is required. This was most effective at reducing early hospital readmissions in mental healthcare (Lepping et al.,2015). Extrapolating findings from such studies to

physical health improvements and developing policy that introduces alternative models of working for all clinical staff caring for individuals with SMI could be considered. New models of care to enable optimum professional contact without the need for hospitalisation might include home visits and community based work; however, reductions in health inequalities would need to involve primary and emergency room care as well as mental healthcare support.

Further work could usefully identify and disentangle substance use among individuals known to have SMI with aim to developing a robust national policy on how they can be managed in acute settings. As reported in Chapter 7, I found that admissions to general hospitals with a primary discharge diagnosis of a mental and behavioural disorder was 30-fold higher for individuals with SMI. Based on the inclusion criteria for the study, these individuals were also known to have contact with their mental health services. A proportion of these admissions are likely to be due to alcohol and/or other substance use disorders which clearly needs to be more effectively highlighted in policy and practice.

The use of resource-intensive hospital services where clear diagnoses were not found has commissioning implications for SMI care. As reported in Chapter 7, I found that more individuals with SMI were assigned to "symptoms, signs and findings, not elsewhere classified- (R00-R99)" and "factors influencing health status and contact with health services- (Z00- Z99). This might indicate inappropriate use of general hospital services due to lack of engagement with other sectors, such as primary care; however, it might also

indicate under-diagnosis. Both lack of engagement and under-diagnosis could result in higher risk of adverse outcomes and might be amenable to more assertive, engaging and accessible models of service delivery. The need for ongoing improvements to databases is one implication to improve the granularity of information on given hospitalisation episodes. Improving awareness among clerical staff who input data that feed into these databases is also key. Education of clinical staff might be another consideration as the coding process begins with the analysis and abstraction of a medical report.

Appendix 1- Symptom descriptions, annotation guidelines and performance metrics

Acknowledgement- Appendix 1 is a document that was produced by BRC informatics team member Jyoti Jyoti. Other colleagues including myself have provided input to its content.

Appendix Table 1: Performance of symptom information extraction applications for individual symptoms

Symptoms	Precision	Recall
Abstract thinking	0.89	1.00
Affect Instability	0.73	0.75
Aggression	0.79	0.96
Agitation	0.97	0.80
Anergia	0.86	1.00
Anhedonia	0.88	0.93
Appetite	0.83	0.87
Blunted Flat Affect	0.94	0.56
Catatonic Syndrome	0.74	0.92
Circumstantial Speech	0.88	0.95
Coherence	0.80	0.90
Concentration	0.88	0.85
Delusion	0.90	1.00
Derailment of Speech	0.88	0.95
Disturbed Sleep	0.77	0.97

Symptoms	Precision	Recall
Echolalia	0.87	0.97
Echopraxia	0.74	0.92
Elation	0.83	1.00
Elevated Mood	0.75	0.93
Emotion Instability	0.70	0.83
Foi	0.97	0.99
Ftd	0.83	0.88
Grandiosity	0.88	0.93
Guilt	0.77	0.95
Hallucination	0.95	0.97
Helpless	0.84	0.99
Hopeless	0.84	0.98
Hostility	0.87	1.00
Immobility	0.82	0.96
Insomnia	0.88	0.93
Irritability	0.82	0.98

Symptoms	Precision	Recall
Low Mood	0.84	0.83
Mannerism	0.66	0.79
Mdma	0.89	1.00
Mood Instability	0.70	0.80
Mutism	0.90	1.00
Paranoia	0.90	1.00
Persecution	0.92	0.99
Perseverance	0.78	0.98
Poor Motivation	0.93	0.77
Posturing	0.83	0.95
Poverty of Speech	0.92	0.97
Poverty of Thought	0.92	0.97
Pressurised Speech	0.77	0.90
Psychomotor	0.83	0.80
Social Withdrawal	0.79	0.88
Stereotype	0.77	0.94

Symptoms	Precision	Recall
Stupor	0.79	0.87
Tangential Speech	0.93	1.00
Tremor	0.95	0.83
Waxy Flexibility	0.87	0.87
Worthless	0.85	0.98

1. Developed GATE Symptoms Applications

1. Anhedonia App

Description

Anhedonia means the inability to experience pleasure from activities usually found enjoyable. Application to identify patients who have experienced anhedonia.

Classes produced: Positive, Negative and Unknown.

Positive mentions were fairly self-explanatory – past or present symptoms included (X had been anhedonic, X has anhedonia etc.)

Negative mentions again were self-explanatory (no anhedonia, no evidence of anhedonia, not anhedonic etc.)

'Unknown' annotations included: i) used in a list, not applying to patient (e.g. typical symptoms include …); ii) uncertain (might have anhedonia, ?anhedonia, possible anhedonia); iii) not clearly present (monitor for anhedonia, anhedonia has improved); iv) listed as potential treatment side-effect; v) vague ('she is not completely anhedonic', 'appears almost anhedonic')

Search Terms:

[Aa]nhedon*

Precision and Recall

Precision = 0.88, Recall = 0.93

2. Anergia App

Description

Application to extract occurrences of Anergia.

Classes produced: Positive, Negative and Unknown.

Positive mentions were fairly self-explanatory – past or present symptoms included (feeling of anergia etc.)

Negative mentions again were self-explanatory (no anergia, no evidence of anergia, no feeling of anergia etc.)

Search Terms

[Aa]nergia

Precision and Recall

P=0.86, R=1.0

3. Energy App

Description:

Application to extract occurrences of energy.

Classes produced: Positive, Negative and Unknown.

Positive mentions were fairly self-explanatory (full of energy etc.)

Negative mentions again were self-explanatory (no energy etc.)

Search Terms

[Ee]nergy

Precision and Recall

P=0.81, R=0.76

4. **Helpless App**

Description:

Application to extract occurrences of helplessness.

Classes produced: Positive, Negative and Unknown.

Positive mentions were fairly self-explanatory (Ideas of helplessness secondary to her physical symptoms present , ideation Compounded by anxiety and a sense of helplessness, hopelessness etc.)

Negative mentions again were self-explanatory (Denies uselessness or helplessness , No thoughts of hopelessness or helplessness etc.)

Search Terms

[Hh]elpless

Precision and Recall

P=0.84, R=0.99

5. **Hopeless App**

Description

Application to extract occurrences of hopelessness.

Classes produced: Positive, Negative and Unknown.

Positive mentions were fairly self-explanatory (feeling very low and hopeless , says feels hopeless etc.)

Negative mentions again were self-explanatory (Denies hopelessness , No thoughts of hopelessness or helplessness etc.)

Unknown mentions again were self-explanatory (unable to say if he feels hopeless etc)

Search Terms

[Hh]opeles

Precision and Recall

P=0.84, R=0.98

6. **Psychomotor App**

Description

Application to extract occurrences of psychomotor.

Classes produced: Positive and Negative.

Positive mentions were fairly self-explanatory (few spontaneous movements-psychomotor retardation etc.)

Negative mentions again were self-explanatory (No psychomotor retardation etc.)

Unknown mentions again were self-explanatory (unable to say if he feels hopeless etc)

Search Terms

[Pp]sychomotor

Precision and Recall

P=0.83, P=0.80

7. Tearful App

Description

Application to extract occurrences of tearful.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: appeared tearful; was tearful (including was XX and tearful; was tearful and YY); became tearful; moments of tearfulness; a bit tearful.

Negative annotations include: not tearful; no tearfulness; denies feeling tearful; no tearful episodes.

‘Unknown’ annotations were mostly ambiguous statements (e.g. less tearful; couldn’t remember being tearful) and statements applying to another person (e.g. mother was tearful) or a person who was not clearly enough the patient.

Search Terms:

[Tt]earful

Precision and Recall:

P=0.91, R=0.99

8. Weight loss App

Description:

Application to extract occurrences of weight loss.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: appeared tearful; was tearful (Significant weight loss, pleased with her weight loss).

Negative annotations include: no weightloss; denies weightloss.

Unknown annotations include: To maintain adequate dietary intake and avoid weight loss, The latter reduced in line with weight loss.

Search Terms:

[Ww]eight [0 to 2 words in between] [Ll]oss

[Ll]oss [0-2 words in between] [Ww]eight

Precision and Recall:

P=0.75, R=0.97

9. Worthless App

Description:

Application to extract occurrences of worthlessness.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: appeared tearful; was tearful (feeling worthless, feels hopeless and worthless).

Negative annotations include: no worthlessness; denies worthlessness etc.

Unknown annotations include: his father had told him that he was worthless, would call them worthless etc

Search Terms:

[Ww]orthless

Precision and Recall:

P=0.85, R= 0.98

10. Appetite App

Description:

Application to extract symptoms of Appetite. Appetite is one of the symptoms of depression. One can experience **lack of appetite** (more common) or **over-eating.**

Classes produced: Positive, Negative and Unknown.

Positive annotations applied to adjectives implying a good or normal appetite: fine; OK; reasonable; alright; preserved; satisfactory. Often described in combination with other symptoms (e.g. sleep and appetite normal; sleep and appetite: both preserved).

Negative annotations applied to adjectives implying a poor/declining appetite: loss of; reduced; decrease in; not so good; diminished; lack of; not great. Also often in combination with other symptoms (poor sleep and appetite; loss of energy and appetite).

'Unknown' annotations include insufficiently informative adjectives: not changed; varies; increased; improving. Also hypothetical mentions, as a potential side effect, as an early warning sign, as a description of a diagnosis (rather than patient experience), describing a relative rather than the patient, 'appetite suppressants'.

From the text – we'd like to extract mentions of normal/good or reduced/poor appetite, rest of the appetite instances are irrelevant (hence un-highlighted).

Normal appetite and reduced appetite will encapsulate the following descriptive terms:

Good or normal appetite	Poor or reduced appetite
Alright	Absent
Eats well	Decreasing
Eating well	Deficit
Excellent	Diminished
Fine	
Fair	Gone down
Good	Loss of
Has appetite	Losing (also loosing)
Healthy	Lost
Intact	Low
Not too bad	Lacking
No problem(s)	Lack of
No concern(s)	Lacks
Not a concern	Less
No issue(s)	Not great

Normal	No
OK(ay)	No interest
Preserved	Not as good
Reasonable	Not very well
Regular	Poor
Stable	Reduced
Satisfactory	Reduction
Steady	Small(er)
Unremarkable	Suppress(ed)
Unimpaired	Suppression
Denies problems with	Worse
Denies issues with	Worsening

Search Terms:

[Aa]ppetite

Precision and Recall:

P=0.83, R=0.87

11. Suicide App

Description:

A non-fatal, self- directed potentially injurious behaviour with intent to die as a result of the behaviour; might not result in injury.

Definition of Suicide Attempt from ICD 10 online (Intentional self-harm: X60 – X84)

X60 - Intentional self-poisoning by and exposure to non-opioid analgesics, antipyretics and anti-rheumatics

X61 - Intentional self-poisoning by and exposure to antiepileptic, sedative-hypnotic, anti parkinsonism and psychotropic drugs, not elsewhere classified

X62 - Intentional self-poisoning by and exposure to narcotics and psychodysleptics [hallucinogens], not elsewhere classified

X63 - Intentional self-poisoning by and exposure to other drugs acting on the autonomic nervous system

X64 - Intentional self-poisoning by and exposure to other and unspecified drugs, medicaments and biological substances

X65 - Intentional self-poisoning by and exposure to alcohol

X66 - Intentional self-poisoning by and exposure to organic solvents and halogenated hydrocarbons and their vapours

X67 - Intentional self-poisoning by and exposure to other gases and vapours

X68 - Intentional self-poisoning by and exposure to pesticides

X69 - Intentional self-poisoning by and exposure to other and unspecified chemicals and noxious substances

X70 - Intentional self-harm by hanging, strangulation and suffocation

X71 - Intentional self-harm by drowning and submersion

X72 - Intentional self-harm by handgun discharge

X73 - Intentional self-harm by rifle, shotgun and larger firearm discharge

X74 - Intentional self-harm by other and unspecified firearm discharge

X75 - Intentional self-harm by explosive material

X76 - Intentional self-harm by smoke, fire and flame

X77 - Intentional self-harm by steam, hot vapours and hot objects

X78 - Intentional self-harm by sharp object

X79 - Intentional self-harm by blunt object

X80 - Intentional self-harm by jumping from a high place

X81 - Intentional self-harm by jumping or lying before moving object

X82 - Intentional self-harm by crashing of motor vehicle

X83 - Intentional self-harm by other specified means

X84 - Intentional self-harm by unspecified means

Search Terms:

asphyxiat*, to shoot herself, to shoot himself, to stab herself, to stab himself, after a suicide attempt, another attempt to end [0 to 1 word in between] life, another suicidal attempt, another suicide attempt, attempt on her life, attempt on his life, attempt to commit suicide, attempt to end her life, attempt to end his life, attempt to kill herself, attempt to kill himself, attempt to suicide, attempt to take her life, attempt to take his life, attempted to commit suicide, attempted to end her life, attempted to end his life, attempted to kill herself, attempted to kill himself, attempted to take her life, attempted to take his life, attempts [1 to 3 words in between] suicide, attempts [0 to 2 words in between] suicide, attempts suicide, depression with suicide attempt, E850, E851,E852,E853,E854,E855, E856, E857,E858,E950,E951,E952,E953,E954,E955,E956,E957,E958,E959,E980,E981,E982,E983,E984,E985,E986, E987,E988,E989,failed suicidal attempt, failed suicide attempt, first attempt at her life, first attempt at his life, following a suicide attempt, following her suicidal attempt, following her suicide attempt, following his suicidal attempt, following his suicide attempt, he jumped off a bridge, his suicide attempt was not successful, history of suicide attempt, in an attempt to kill herself, in an attempt to kill himself, leapt in front of motor, leapt in front of moving, leapt in front of traffic, leapt in front of vehicle, next time he attempts suicide, next time she attempts suicide, previous [1 word in between] attempt by, previous [1 word in between] attempt to, previous [1 word in between] attempt to take [1 word in between] life, previous [1 word in between] attempt to take life, previous [1 word in between] attempt by, previous [1 word in between] attempt to, previous [1 word in between] attempt to take [1 word in

between] life, previous [1 word in between]attempt to take [1 word in between] life, previous [1 word in between] attempt to take life, previous attempt by, previous attempt to, previous attempt to take [1 word in between] life, previous attempt to take [1 word in between] life, previous attempt to take life, she jumped off a bridge, shot herself, shot himself, stabbed herself, stabbed himself, suicidal [1 to 3 words in between]attempt, suicidal [1-3 words in between]attempt, suicidal [1 to 2 words in between] attempt, suicidal [1to 2 words in between] attempt, suicidal [1 word in between] attempt, suicidal [1 to 3 words in between]attempt, suicidal [1 to 3 words in between] attempt, suicidal [1 to 3 words in between] attempt, suicidal attempt due to depression, suicidal attempt was not successful, suicidal behaviour, suicidal gesture, suicide [1 to 3 words in between] attempt, suicide [1 to 2 words in between] attempt, suicide [1 to 3 words in between] attempt, suicide [1 to 2 words in between] attempt, suicide [1 word in between] attempt, suicide [1 to 3 words in between] attempt, suicide [1 to 3 words in between] attempt, suicide [1 to 2 words in between] attempt, suicide [0 to 2 words in between] attempt, suicide attempt age, suicide attempt due to depression, suicide attempt failed, suicide attempt was not successful, suicide behaviour, suicide behaviour, suicide gesture, took firearm to self, took gun to self, took handgun to self, took knife to self, took rifle to self, tried to end her life, tried to end his life, tried to kill herself, tried to kill himself, tried to take her life, tried to take his life, walked in front of cars, walked in front of motor, walked in front of moving, walked in front of traffic, walked in front of vehicles, X60, X61, X62, X63, X64, X65, X66, X67, X68, X69, X70, X71, X72, X73, X74, X75, X76, X77, X78, X79, X80, X81, X82, X83, X84, Y87.0, Z72.8, Z91.5, crash*, cutting her wrist, cutting his wrist, drown*, firearm*, handgun, hang*, impulsive O/d, ingested*, jumped off, overdose, overdosing, poison*, putting

plastic bag over her head, putting plastic bag over his head, stabbing himself, stabbing herself, attempted to hang himself, attempted to hang herself, started to cut, starting to cut, strangl*, strangul*, submerg*, suffocate herself, suffocate himself, suffocated herself, suffocated himself, taken an O/D, taken an OD, took [8 - 100] tablets, took an OD, took O/D, tried to cut, walked into traffic, wrist cutting

Important Notes Before Coding

Any mention where the patient has undergone/experience intentional harm to self (physical harmful action directed at self) should be counted as suicide attempt. Literal mentions of "self-harm" are ignored because of danger of over-including DSH events which are different to a suicide attempt.

Classification/ Coding Rules

Rules followed when coding Suicide Attempt in CRIS Notes:

Code Suicide attempt as **positive** if it meets the description of a Suicide attempt (See "Definition of a Suicide Attempt" above)

Code Suicide attempt as **negative** if patient or clinical staff are denying or negating the suicide attempt (e.g. "she never attempted suicide")

Code Suicide attempt as **negative** if patient or clinical staff are referring to family member who attempted suicide (e.g. "mum attempted suicide")

Code Suicide attempt as **negative** where mention of suicide attempt is irrelevant (e.g. "in theory suicide attempts are harsh")

Code Suicide attempt as negative where mention of suicide attempt is confusing or uncomprehensive

Coding Suicide attempt as **negative** where mention of suicide attempt a literal mention of "self-harm"

Precision and Recall:

P=0.86, R=0.97

12. Abstract Thinking App

Description:

Application to extract occurences of concrete thinking.

Classes produced: Positive, Negative and Unknown.

Positive annotations includes: Occurrences of Concrete thinking or concrete cognition.

Unknown indicates- Concrete answers about one particular area.

Negative annotations includes: no evidence of concrete thinking.

Search Terms:

[Cc]oncrete

Precision and Recall:

P=0.89, R=0.60

Comments/ Coding Rules:

Very few occurrences of the negative class (e.g. No evidence of concrete thinking)

Concrete thinking or concrete cognition is the main positive occurrences.

Concrete answers about one particular area classed as 'Unknown'

13. Apathy App

Description:

Application to extract the presence of apathy, developed as part of the Negative Symptoms of Schizophrenia project.

Classes produced: Positive, Negative and Unknown.

Positive mentions include any indication that apathy was being reported as a symptom: e.g. continues to demonstrate A; symptoms include A; he is withdrawn, attributable to A; his apathy … ; some degree of A noted; presentation with A; his report of A; given patient's level of A. Most apathy statements were found to be accompanied by 'negative symptoms' (i.e. rather than depressive).

Very few negative mentions found. Usual statements (denied A; no evidence of A etc.)

'Unknown' annotations include: apathy mentioned as a hypothetical cause of something else (e.g. inactivity) rather than described as being present; apathy mentioned as a possibility in the future (e.g. may develop A), or as a possible side effect of medication (rather than actually present), or as an early warning sign. Also *apathy* found in quite a few names.

Search Terms:

[Aa]path

Precision and Recall:

P=0.86, R=1.0

14. Circumstantial App

Description:

Application to extract occurrences where circumstantiality is apparent.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: Signs of over inclusiveness and circumstantially, Loose associations and circumstantiality, circumstantial in nature etc.

Negative mentions include: No signs of Circumstantially, no evidence of circumstantial.

Unknown mentions include: Circumstantial mentioned as a hypothetical cause of something else.

Search Terms:

[Cc]ircumstance*

Precision and Recall:

P=0.88, R=0.95

15. Delusion App

Description:

Application to extract occurrences where delusions are apparent.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: paranoid delusions; continued to express delusional ideas of the nature etc.

Negative mentions include: No delusion, denied delusion etc.

Unknown mentions include: Present or Absent of Delusion is not clear.

Search Terms:

[Dd]elusion

Precision and Recall:

P=0.90, R=1.0

16. Disturbed Sleep App

Description:

Application to extract the occurrences of disturbed sleep.

Classes produced: Positive, Negative and Unknown.

Positive mentions include the following sorts of phrases: complains of poor sleep, poor sleep, sleep disturbed, sleep difficulty, sleeping poorly, not sleeping very well, cannot sleep, sleep pattern poor, difficulties with sleep

Negative mentions include the following sorts of phrases: no sleep disturbance, sleeping without problems, no significant problems with sleep, sleep and [appetite] undisturbed, sleeping well, denied [xxx] and sleep disturbance.

'Unknown' annotations include the following: call [nurse] if sleep disturbance, sleeping tablets, hoping to sleep, encourage to sleep, medication to help sleep, has been asleep all day, less disturbed sleep today, sleep disturbance as an early warning sign, sleeping habits present no problem, xxx might lead to sleep problems, settled down to sleep.

Search Terms:

Not sleep, sleep not, poor [0-5 words in between] sleep, sleep poor, sleep [0-5 words in between] poor, sleep [0-5] interrupt, interrupt [0-5 words in between] sleep, disturb [0-5 words in between] sleep, sleep [0-5 words in between] disturb, inadequate [0-5 words in between] sleep, sleep [0-5 words in between] inadequate, disorder [0-5 words in between]sleep, sleep [0-5 words in between] disorder, prevent [0-5 words in between] sleep, sleep [0-5 words in between] prevent, stop [0-5 words in between] sleep, sleep [0-5 words in between] stop, sleep [0-5 words in between] problem, problem [0-5 words in between] sleep

Precision and Recall:

P=0.77, R=0.97

17. Insomnia App

Description:

Application to extract the occurrence of Insomnia.

Classes produced - Positive, Negative and Unknown.

Positive mentions included any insomnia described (including initial insomnia, middle insomnia), any assumed application to the patient - 'the insomnia', complaining of insomnia, taking X for insomnia, contributes to her insomnia, problems with insomnia, this has resulted in insomnia, this will address his insomnia. Negative mentions again were self-explanatory (no anhedonia, no evidence of anhedonia, not anhedonic etc.)

'Unknown' annotations included: i) used in a list, not applying to patient (e.g. typical symptoms include …); ii) uncertain (might have anhedonia, ?anhedonia, possible anhedonia); iii) not clearly present (monitor for anhedonia, anhedonia has improved); iv) listed as potential treatment side-effect; v) vague ('she is not completely anhedonic', 'appears almost anhedonic')

Search Terms:

[Ii]nsom*

Precision and Recall:

P=0.88, R=0.93

18. Irritability App

Description:

Application to extract the occurrence of Irritability.

Classes produced - Positive, Negative and Unknown.

Positive mentions included both observed and self-reported/subjective irritability, 'can be irritable', 'became irritable', 'appeared irritable', 'complained of feeling irritable'.

Negative mentions included no evidence of, no longer, no sign of (again, both observed and subjective)

'Unknown' annotations included unclear statements, irritable bowel syndrome, 'becomes irritable when unwell', early warning sign, [relative] gets irritable, listed as potential side effect / withdrawal symptom, 'can be irritable if …' [NB some ambiguity with positive 'can be' mentions, although linked here with the 'if' qualifier], 'less irritable'.

Search Terms:

[Ii]rritabl*

Precision and Recall:

P=0.82, R=0.98

19. Poor Motivation App

Description:

Application to establish the level of motivation, developed as part of the Negative Symptoms of Schizophrenia project.

Classes produced - Positive, Negative and Unknown.

Positive annotations applied to any statement describing motivation as poor or diminished: lack of; lacking in; does not have the M to …; finds it difficult to M herself; poor M; is not motivated to …; a motivation; zero motivation; unmotivated; reduced M.

Negative annotations applied to any statement describing motivation as good or normal: denies difficulties with M; good M; appears motivated to …; highly motivated.

No 'unknown' annotations. NB annotations above were applied regardless of what the motivation was focused on (i.e. including both general statements of good/poor motivation, and also task-specific statements of motivation to do XX or YY)

Search Terms:

[Mm]otivat

Precision and Recall:

P=0.93, R=0.77

20. Poverty of thought App

Description:

Application to extract whether poverty of thought is present.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: poverty of thought was very striking, evidence of poverty of thought etc.

Negative mentions include: no poverty of thought, no evidence of poverty of thought etc.

Unknown mentions include: Present or Absent of Poverty of thought is not clear.

Search Terms:

[Pp]overt [Oo]f [Tt]hough

Precision and Recall:

P=0.92, R=0.97

21. Aggression App

Description:

Application to identify mentions of aggression.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: Reported to be quite aggressive towards …, Violence and aggression, requires continued management and continues to reduce in terms of incidents etc.

Negative mentions include: no aggression, no evidence of aggression etc.

Unknown mentions include: annotations included unclear statements.

Search Terms:

[Aa]ggress

Precision and Recall:

P=0.79, R=0.96

22. Agitation App

Description:

Application to identify mentions of agitation.

Classes produced: Positive, Negative, and Unknown.

Positive mentions include: very agitated at present, he was agitated, He was initially calm but then became agitated and started staring and pointing at me towards.. etc.

Negative mentions include: did not seem distracted or agitated, not agitated, no evidence of agitation.

Unknown mentions include: annotations include unclear statements.

Search Terms:

[Aa]gitat*

Precision and Recall:

P=0.79, R=0.96

23. Arousal App

Description:

Application to identify mentions of arousal.

Classes produced: Positive, Negative and Unknown.

Positive mentions include: became irritable and aroused on occasion, ZZZZZ became very aroused when talking.., increasingly aroused and hostile towards staff etc.

Negative mentions include: denies arousal at the hostel, he is not currently aroused or agitated etc.

Unknown mentions include: annotations include unclear statement.

Search Terms:

[Aa]rous*

Precision and Recall:

P=0.87, R=1.0

24. Blunted Flat Affect App

Description:

Application to extract the presence of Blunted Flat Affect.

Classes produced: Positive and Negative.

Positive annotations include: His affect remains very blunted, Objectively flattened affect, State ZZZZZ continues to appear flat in affect etc.

Negative annotations include: incongruent affect, Stable affect etc.

Search Terms:

[Aa]ffect

[Bb]lunt [0 to 2 words in between] [Aa]ffect

[Ff]lat [0 to 2 words in between] [Aa]ffect

[Rr]estrict [0 to 2 words in between [Aa]ffect

[Aa]ffect [0 to 2 words in between] [Bb]lunt

[Aa]ffect [0 to 2 words in between] [Ff]lat

[Aa]ffect [0 to 2 words in between] [Rr]estrict

Precision and Recall:

P=0.94, R = 0.56

25. Catatonia Catalepsy App

Description:

Application to extract occurrences where catalepsy is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: to long-standing narcolepsy and catalepsy, Of note, today he presented catalepsia etc.

Negative annotations include: No catalepsy, no evidence of catalepsia etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Cc]atalep*

Precision and Recall:

P=0.74, R=0.92

26. Echolalia App

Description:

Application to extract occurrences where echolalia is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: No neologisms, but repeated what I said almost like echolalis; intermittent echolalia etc.

Negative annotations include: no echolalia, no evidence of echolalia, cannot remember any echolalia or stereotyped utterances etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Ee]chola*

Precision and Recall:

P=0.87, R=0.97

27. Echopraxia App

Description:

Application to extract occurrences where echopraxia is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: displayed a lot of echolalia and echopraxia, some of which may be triggers for ZZZZZ 's echopraxia etc

Negative annotations include: He did not elicit any echolalia, echopraxia or coprolalia, There were no repetitive behaviours, echolalia, echopraxia etc.

Unknown annotations include: annotations include unclear statements such as Mrs ZZZZZ to review her mimicry behaviours with a view to providing another opinion

regarding whether they are best understood as tics (echopraxia), or whether they may be compulsive rituals

Search Terms:

[Ee]chopra*

Precision and Recall:

P=0.74, R=0.92

28. Immobility App

Description:

Application to extract occurrences where immobility is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: on occasions she feels miserable and low due to her immobility, Recent immobility is a sudden change over the last week or so etc.

Negative annotations include: immobile swelling approx 5cm above ankle on anterior shin, no immobility etc.

Unknown annotations include: annotations include unclear statements such as to have injections Mon and will be immobile early next week as a result etc.

Search Terms:

[Ii]mmobil

Precision and Recall:

P=0.82, R=0.96

29. Mannerism App

Description:

Application to extract occurrences where mannerism is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: covering his face with a scarf, laughing inappropriately, and talking to himself, prating, and exhibiting frequent mannerisms, the episodes/repetitive mannerisms could be attributed to excitement and an attention seeking component etc .

Negative annotations include: Somewhat dramatic in his mannerisms, using hand gestures to emphasise his speech, no mannerisms apart from moving his mouth etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Mm]annerism

Precision and Recall:

P=0.66, R=0.79

30. Perseverance App

Description:

Application to extract occurrences where perseverance is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: which might manifest themselves in poor organisation skills and a tendency to perseverate; Thought: Contents - perseverates re hospitalisation in … etc

Negative annotations include: no evidence of any flights of ideas, perseveration or circumstantialities noted; indicated that she was happy to persevere with her current regime etc

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Pp]ersev*

Precision and Recall:

P=0.78, R=0.98

31. Posturing App

Description:

Application to extract occurrences where posturing is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: he remained in an immobile posture, Stillposturing in the corridor or sitting in a dark interview room etc.

Negative annotations include: No abnormal movement or posture, He maintains an aggresive posture while he paces the unit etc.

Unknown annotations include: annotations include unclear statements such as Her posture changed and she looked quite tearful etc.

Search Terms:

[Pp]ostur

Precision and Recall:

P=0.83, R=0.95

32. Rigid App

Description:

Application to extract occurrences where rigidity is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: he did have some rigidity, rigid body movements etc.

Negative annotations include: No rigidity noted, No rigidity in all four limbs etc.

Unknown annotations include: annotations include unclear statements such as He stated that rigid time slot would not be suitable, Her impression impression was a damaged man with rigid patterns of self-sabotage etc.

Search Terms:

[Rr]igid

Precision and Recall:

P=0.82, R=0.73

33. Stereotype App

Description:

Application to extract occurrences where stereotype is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: he developed catatonic features - mannerisms and stereotypes; unable to sit still during interview with stereotypic and purposeless actions etc.

Negative annotations include: no stereotypic behaviour, no evidence of stereotype etc.

Unknown annotations include: annotations include unclear statements such as he mentions that he was not oblivious to the stereotype of children.

Search Terms:

[Ss]tereotype*

Precision and Recall:

P=0.77, R=0.94

34. Stupor App

Description:

Application to extract occurrences where stupor is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: Admitted with depressive stupor, Tonight ZZZZZ presented in a psychotic stupor etc.

Negative annotations include: not in the state of stupor /confusional state, not suggestive of depressive stupor etc.

Unknown annotations include: annotations include unclear statements such as On a few occasions he has resorted to drinking or smoking into a stupor in order to knock him self out etc.

Search Terms:

[Ss]tupor

Precision and Recall:

P=0.79, R=0.87

35. Waxy Flexibility App

Description:

Waxy flexibility is a psychomotor symptom of catatonia as associated with schizophrenia, bipolar disorder, or other mental disorders which leads to a decreased response to stimuli and a tendency to remain in an immobile posture.

Application to extract occurrence where waxy flexibility is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: she presents as catatonic with waxy flexibility, exhibiting waxy flexibility etc.

Negative annotations include: His right pre-tibial region was swollen and waxy and slightly pink, no waxy flexibility etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Ww]axy

Precision and Recall:

P=0.87, R=0.87

36. Circumstantial App

Description:

Application to extract occurrences where circumstantial is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: loose associations a loss of goal, overinclusiveness and circumstantially; He had thought derailment and circumstantially etc.

Negative annotations include: The five Risk items (R) are concerned with forecasting the future social, living and treatment circumstances etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Cc] ircumstan*

Precision and Recall:

P=0.88, R=0.95

37. Coherence App

Description:

Application to extract occurrences where coherence is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: he was rather incoherent in his speech , At the start of the interview his talk was incoherent etc.

Negative annotations include: he could not hold a coherent conversation, she was worried about him as he was sounding confused and incoherent etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Cc] ircumstan*

Precision and Recall:

P=0.80, R=0.90

38. Derailment of Speech App

Description:

Application to extract occurrences where derailment of speech is present.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: he derailed frequently, There was evidence of flight of ideas and thought derailment in his language etc.

Negative annotations include: The thought stream is normal as he uses sentences in consequences, with no derailment, Erratic compliance can further derail her stability etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Dd]erail*

Precision and Recall:

P=0.85, R=1.0

39. Elation App

Description:

Application to extract occurrences where elation is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: mildly elated in mood, elated in mood on return from leave, She appeared elated and aroused etc.

Negative annotations include: ZZZZZ was coherant and more optimistic/apirational than elated throughout the conversation, no elated behaviour etc.

Unknown annotations include: annotations include unclear statements such as To others - no history of harm to others but in his elated state there is a risk of accidental harm to others as is pushing boundaries etc.

Search Terms:

[Ee]lat*

Precision and Recall:

P=0.83, R=1.0

40. Elevated Mood App

Description:

Application to extract occurrences where elevated is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: ZZZZZ 's paranoia and elevated mood have been resulting in conflict with others on the unit; Elevated mood, lability. Self medication with drugs;

Negative annotations include: No overt evidence of mood elevation; No period of elevated mood/irritability etc.

Unknown annotations include: annotations include unclear statements such as He was visited by his girlfriend and that elevated his mood etc.

Search Terms:

[Mm]ood

Precision and Recall:

P=0.75, R=0.93

41. Emotional Withdraw App

Description:

Application to extract occurrences of emotional withdrawal, developed as part of the Negative Symptoms of Schizophrenia project.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: Her mother described has becoming withdrawn, not communicating; . Sabri was visible on the ward appears withdrawn etc.

Negative annotations include: He was not withdrawn; He was not observed to be sad or withdrawn during the assessment etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Ww]ithdraw

Precision and Recall:

P=0.75, R=0.93

42. Flight of Idea App (foi)

Description:

Application to extract occurrences where flight of ideas is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: Mrs ZZZZZ was very elated with by marked flights of ideas; marked pressure of speech associated with flights of ideas etc.

Negative annotations include: no evidence of flight of ideas, no flight of idea etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Ff]light [Oo]f [Ii]dea

Precision and Recall:

P=0.97, R=0.99

43. Formal thought disorder(ftd)

Description:

Application to extract occurrences where formal thought disorder is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: deteriorating into a more thought disordered state with outbursts of aggression; There was always a degree thought disorder etc.

Negative annotations include: Thoughts: No FTD; No signs of FTD.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Ff][Tt][Dd]

[Ff]ormal [Tt]ought [Dd]isorder

Precision and Recall:

P=0.83, R=0.88

44. Grandiosity App

Description:

Application to extract occurrences where grandiosity is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: ZZZZZ was wearing slippers and was animated elated and grandiose, few grandiose statements regarding having been 'brought up with royalty' etc.

Negative annotations include: No evidence of grandiose of delusions in the content of his speech, No evidence of grandiose ideas etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Gg]randios

Precision and Recall:

P=0.88, R=0.93

45. Hallucination App

Description:

Application to extract occurrences where hallucination is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: recalled experiencing auditory hallucinations, visual distortions and auditory hallucinations etc.

Negative annotations include: No evidence of responding to any hallucinations or delusions, Not objectively hallucinated etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Hh]allucinat*

[Aa]udit [0 to 3 words in between] [Hh]allucinat*

[Vv]isual [0 to 3 words in between] [Hh]allucinat*

[Oo]lfact [0 to 3 words in between] [Hh]allucinat*

[Tt]actil [0 to 3 words in between] [Hh]allucinat*

[Tt]hird person [0 to 3 words in between] [Hh]allucinat*

[Ff]irst person [0 to 3 words in between] [Hh]allucinat*

3rd person [0 to 3 words in between] [Hh]allucinat*

1st person [0 to 3 words in between] [Hh]allucinat*

[Hh]allucinat* [0 to 3 words in between] [Aa]udit

[Hh]allucinat* [0 to to 3 words in between] [Vv]isual

[Hh]allucinat* [0 to 3 words in between] [Oo]lfact

[Hh]allucinat* [0 to 3 words in between] [Tt]actil

[Hh]allucinat* [0 to 3 words in between] [Tt]hird person

[Hh]allucinat* [0 to 3 words in between] [Ff]irst person

[Hh]allucinat* [0 to 3 words in between] [3rd person]

[Hh]allucinat* [0 to 3 words in between [1st person]

Precision and Recall:

P=95, R=0.97

46. Hostility App

Description:

Application to extract occurrences where hostility is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: increased hostility and paranoia, she presented as hostile to the nurses etc.

Negative annotations include: no hostile, Denied any feelings of hostility etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Hh]ostil*

Precision and Recall:

P=0.87, R=1.0

47. Mutism App

Description:

Application to extract occurrences where mutism is apparent.

Classes produced: Positive and Unknown.

Positive annotations include: She has periods of 'mutism' Issues with landlord, he did not respond any further and remained mute etc.

Unknown annotations include: annotations include unclear statements such as that her mother is mute, Body language muted etc.

Search Terms:

[Mm]ute, [Mm]utism

Precision and Recall:

P=0.90, R=1.0

48. Paranoia App

Description:

Application to extract occurrences where paranoia is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: vague paranoid ideation, cause him to feel paranoid etc.

Negative annotations include: denied any paranoid, no paranoid feeling etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Pp]aranoi*

Precision and Recall:

P=0.90, R=1.0

49. Persecutory Ideation App

Description:

Application to extract occurrences where persecution is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: she was having delusions of persecution, She suffered persecutory delusion etc.

Negative annotations include: Denies persecutory delusions, He denied any worries of persecution etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Pp]ersecu*

Precision and Recall:

P=0.92, R=0.99

50. Poverty of Speech App

Description:

Application to identify poverty of speech, developed as part of the Negative Symptoms of Schizophrenia project.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: He continues to display negative symptoms including blunting of affect, poverty of speech; He does have negative symptoms in the form of poverty of speech etc.

Negative annotations include: No poverty of speech, There was no pressure or poverty of speech etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Pp]overty [0 to 2 words in between] [Ss]peech

[Ii]mpoverish [0 to 2 words in between] [Ss]peech

[Ss]peech [0 to 2 words in between] [Pp]overty

[Ss]peech [0 to 2 words in between] [Ii]mpoverish

Precision and Recall:

P=0.92, R=0.97

51. Pressured Speech App

Description:

Application to extract occurrences where pressured speech is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: Talking with pressured speech; He had pressured speech etc.

Negative annotations include: no pressure of speech, Her speech appeared normal in rate, volume and pressure etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Ss]peech

Precision and Recall:

P=0.85, R=0.98

52. Social Withdrawal App

Description:

Application to identify social withdraw.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: She is withdrawn socially from friends and family, Mr ZZZZZ became very isolated and socially withdrawn etc.

Negative annotations include: not being socially withdrawn, no evidence of socially withdrawn etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Ww]ithdraw*

Precision and Recall:

P=0.79, R=0.88

53. Tangential Speech App

Description:

Application to extract occurrences where tangential speech is apparent.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: He was very tangential lacked goal directed thinking, There was evidence of tangential etc.

Negative annotations include: no evidence of formal thought disorder or tangentiality of thoughts; however there was no overt tangentiality or loosening of associations etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Tt]angent*

Precision and Recall:

P=0.93, R=1.0

54. Thought Block App

Description:

Application to extract whether thought block is present.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: Showed some thought block; thought block and paucity of thought etc.

Negative annotations include: Denies problems with thought insertion; No formal thought disorder etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Tt]hought block

Precision and Recall:

P=0.82, R=0.95

55. Guilt App

Description:

Application to identifies when patients have occurrences of guilt or not.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: she then feels guilty/angry towards mum; Being hungry is easier to deal with than feeling guilty etc.

Negative annotations include: no feeling of guilt; Denies feeling hopeless or guilty etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Gg]uil

Precision and Recall:

P=0.77, R=0.95

56. Concentration App

Description:

Application to identifies when patients have poor or good concentration.

Classes produced: Positive, Negative and Unknown.

Positive annotations include: My concentration is still poor, she found it difficult to concentrate etc.

Negative annotations include: Good attention and concentration, Participating well and able to concentrate on activities etc.

Unknown annotations include: annotations include unclear statements.

Search Terms:

[Cc]oncentrat

Precision and Recall:

P=0.68, R=0.80

57. Low Mood App

Description:

Application to extract occurrences of low mood.

Classes produced: Positive and Negative.

Positive annotations: applied to any description of 'low' applied to mood: describes mood as generally low; appears a bit low in mood; experiencing signs of anxiety and low mood; feels low in mood; subjectively – low in mood; remains low in mood; difficulties with low mood; complaining of … low mood …; his recent misery and low mood; suffering from low mood.

Negative annotations: applied to adjectives describing mood as: normal; good; euthymic; alright; reactive.

No 'unknown' annotations: NB no attempt to distinguish the above annotations in terms of whether subjective or objective mood was sought – just focusing on words describing 'mood'.

Search Terms:

[Mm]ood

Precision and Recall:

P=0.84, R=0.83

Appendix 2- Additional table for Chapter 6

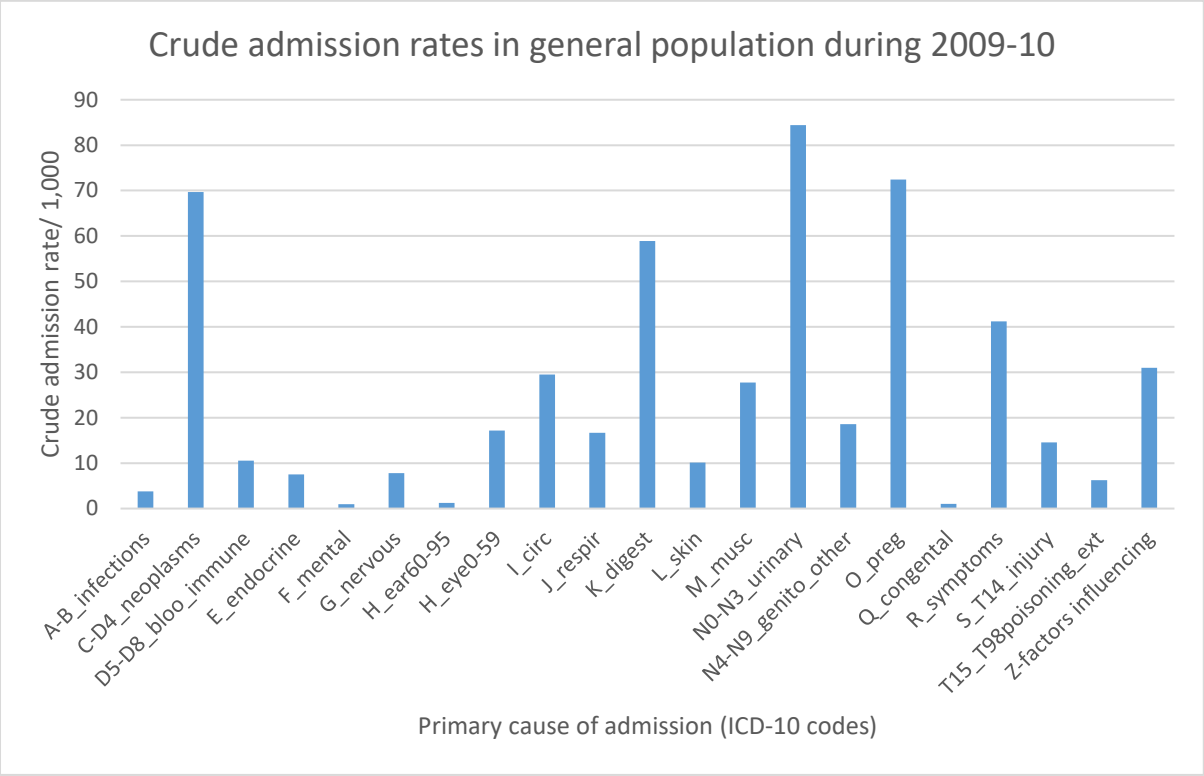
Appendix 2 Table 1- Template abridged life table that was used for life expectancy calculations (using Chiang 1984 method)

Chiang method (II) - Used by ONS for calculation of subnational life expectancy at birth results and 95% confidence limits.

	x	n	a_x	*pop*	*death*	M_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x	$var(q_x)$	$l_x^2[(1-a_x)n_1+e_{x+1}]^2var(q_x)^2$	$sum(l_x^2[(1-a_x)n_1+e_{x+1}]^2var(q_x)^2)$	*Var(ex)*	*SE*	Lower 95% confidence interval	Upper 95% confidence interval
<1	0	1	0.1	2,533	20	0.007896	0.0078401	0.992160	100000	784	99294	7198691	71.99	0.000003049	160076312	842044805	0.08420	0.2902	71.42	72.56
1-4	1	4	0.5	11,130	1	0.000090	0.0003593	0.999641	99216	36	396793	7099397	71.55	0.000000129	6151051	681968493	0.06928	0.2632		
5-9	5	5	0.5	15,519	2	0.000129	0.0006442	0.999356	99180	64	495742	6702604	67.58	0.000000207	8649449	675817442	0.06870	0.2621		
10-14	10	5	0.5	16,409	4	0.000244	0.0012181	0.998782	99116	121	495280	6206862	62.62	0.000000370	13188429	667167993	0.06791	0.2606		
15-19	15	5	0.5	16,133	9	0.000558	0.0027854	0.997215	98996	276	494289	5711582	57.70	0.000000860	25810071	653979564	0.06673	0.2583		
20-24	20	5	0.5	21,482	10	0.000466	0.0023248	0.997675	98720	230	493026	5217293	52.85	0.000000539	13384127	628169493	0.06446	0.2539		
25-29	25	5	0.5	15,997	22	0.001375	0.0068527	0.993147	98490	675	490765	4724267	47.97	0.000002120	43098955	614785366	0.06338	0.2517		
30-34	30	5	0.5	16,026	35	0.002184	0.0108605	0.989140	97816	1062	486422	4233501	43.28	0.000003333	54211405	571686411	0.05975	0.2444		
35-39	35	5	0.5	19,800	34	0.001717	0.0085492	0.991451	96753	827	481698	3747080	38.73	0.000002131	26639208	517475006	0.05528	0.2351		
40-44	40	5	0.5	16,076	39	0.002426	0.0120568	0.987943	95926	1157	476739	3265381	34.04	0.000003682	34536469	490835799	0.05334	0.2310		
45-49	45	5	0.5	13,404	59	0.004402	0.0217688	0.978231	94770	2063	468690	2788642	29.43	0.000007857	53461530	456299330	0.05081	0.2254		
50-54	50	5	0.5	13,027	108	0.008290	0.0406107	0.959389	92706	3765	454120	2319953	25.02	0.000014650	69406482	402837800	0.04687	0.2165		
55-59	55	5	0.5	10,051	136	0.013531	0.0654412	0.934559	88942	5820	430157	1865832	20.98	0.000029429	91009253	333431318	0.04215	0.2053		
60-64	60	5	0.5	10,220	176	0.017221	0.0825516	0.917448	83121	6862	398451	1435675	17.27	0.000035524	63630177	242422065	0.03509	0.1873		
65-69	65	5	0.5	9,190	320	0.034820	0.1601602	0.839840	76259	12214	350763	1037224	13.60	0.000067322	68405970	178791888	0.03074	0.1753		
70-74	70	5	0.5	7,427	445	0.059917	0.2605539	0.739446	64046	16687	278510	686461	10.72	0.000112808	57157515	110385918	0.02691	0.1640		
75-79	75	5	0.5	5,231	414	0.079144	0.3303543	0.669646	47358	15645	197679	407951	8.61	0.000176524	33004917	53228403	0.02373	0.1541		
80-84	80	5	0.5	2,884	355	0.123093	0.4706350	0.529365	31713	14925	121253	210272	6.63	0.000330290	20223485	20223485	0.02011	0.1418		
85+	85	11	0.5	1,840	347	0.188587	1.0000000	0.000000	16788	16788	89019	89019	5.30	0.000000000	0	0	0.00000	0.0000		

Appendix 3- Additional figure for Chapter 7

Appendix 3 Figure 1- Figure showing-crude admission rates/ 1,000 persons in SLaM catchment area during 2009-2010



Appendix 4- Additional tables for Chapter 8

Appendix 4 Table 1: Multivariable ordinal logistic regression analysis of factors associated with negative symptoms in patients with schizophrenia

Factor	Group	Number in sample	CRIS-NSS mean score (range 0-10)	Standard Deviation	Unadjusted		Adjusted model (n=7676)*	
					Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value
Age (years)	16-19	203	1.28	2.15	0.30 (0.22-0.40)	<0.001	0.46 (0.33-0.64)	<0.001
	20-29	1337	2.45	2.64	Reference		Reference	
	30-39	1775	1.97	2.21	0.77 (0.67-0.87)	<0.001	0.78 (0.68-0.89)	<0.001
	40-49	1983	1.66	1.95	0.62 (0.55-0.71)	<0.001	0.62 (0.55-0.71)	<0.001
	50-59	1137	1.43	1.79	0.53 (0.46-0.61)	<0.001	0.52 (0.45-0.61)	<0.001
	60-69	654	1.07	1.53	0.37 (0.31-0.44)	<0.001	0.36 (0.30-0.44)	<0.001
	70+	589	0.72	1.19	0.25 (0.21-0.30)	<0.001	0.25 (0.20-0.31)	<0.001
Gender	Male	4592	1.90	2.22	Reference		Reference	
	Female	3084	1.41	1.91	0.66 (0.60-0.71)	<0.001	0.79 (0.72-0.86)	<0.001
Marital status	Single	5795	1.88	2.20	Reference		Reference	
(most recent)	Married/cohabiting	785	1.16	1.61	0.57 (0.50-0.66)	<0.001	0.74 (0.64-0.86)	<0.001

Factor	Group	Number in	CRIS-NSS mean	Standard	Unadjusted		Adjusted model (n=7676)*	
	Divorced/separated	776	1.35	1.86	0.64 (0.56-0.74)	<0.001	0.86 (0.74-0.99)	0.040
	Widowed	208	0.85	1.39	0.40 (0.31-0.52)	<0.001	0.81 (0.60-1.09)	0.160
Employment (most recent)	Unemployed	4956	2.03	2.23	Reference		Reference	
	Employed	341	1.51	1.87	0.68 (0.56-0.83)	<0.001	0.64 (0.52-0.78)	<0.001
	In education	311	1.64	2.10	0.70 (0.57-0.86)	0.001	0.78 (0.63-0.97)	0.026
	Retired	7	0.57	0.79	0.33 (0.08-1.29)	0.110	0.68 (0.16-2.87)	0.599
ADL impairment	Absent	4700	1.73	2.09	Reference		Reference	
	Present	2283	1.97	2.23	1.21 (1.11-1.33)	<0.001	1.35 (1.22-1.49)	<0.001
Social impairment	Absent	4432	1.76	2.10	Reference		Reference	
	Present	2533	1.88	2.20	1.07 (0.98-1.17)	0.108	0.95 (0.86-1.05)	0.292
Delusions / hallucinations	Absent	3904	1.77	2.17	Reference		Reference	
	Present	3077	1.85	2.11	1.14 (1.05-1.24)	0.003	1.19 (1.09-1.30)	<0.001
Depression	Absent	4976	1.90	2.16	Reference		Reference	
	Present	2014	1.59	2.08	0.71 (0.65-0.79)	<0.001	0.69 (0.62-0.76)	<0.001

Appendix 4 Table 2: Multivariable binary logistic regression analysis of factors associated with negative symptoms in patients with schizophrenia including cases with full covariate data only

Factor	Group	Number in sample	Prevalence of two or more negative symptoms (%)	Association with two or more negative symptoms: odds ratio (95% CI), p-value			
				Unadjusted		Adjusted model (n=5316)*	
Age (years)	16-19	203	27.6	0.35 (0.25-0.49)	<0.001	0.44 (0.29-0.68)	<0.001
	20-29	1337	52.0	Reference		Reference	
	30-39	1775	47.0	0.82 (0.71-0.94)	0.006	0.80 (0.67-0.95)	0.012
	40-49	1983	42.6	0.69 (0.60-0.79)	<0.001	0.64 (0.54-0.76)	<0.001
	50-59	1137	37.2	0.55 (0.47-0.64)	<0.001	0.45 (0.37-0.55)	<0.001

Factor	Group	Number in	Prevalence of two or more	Association with two or more negative symptoms: odds ratio (95% CI), p-			
	60-69	654	29.1	0.38 (0.31-0.46)	<0.001	0.32 (0.25-0.41)	<0.001
	70+	589	18.0	0.20 (0.16-0.26)	<0.001	0.13 (0.09-0.18)	<0.001
Gender	Male	4592	45.3	Reference		Reference	
	Female	3083	34.7	0.64 (0.59-0.71)	<0.001	0.74 (0.65-0.83)	<0.001
Marital status	Single	5795	44.6	Reference		Reference	
(most recent)	Married/cohabiting	785	31.6	0.57 (0.49-0.67)	<0.001	0.77 (0.63-0.94)	0.012
	Divorced/separated	776	33.4	0.62 (0.53-0.73)	<0.001	0.91 (0.75-1.12)	0.376

Factor	Group	Number in	Prevalence of two or more	Association with two or more negative symptoms: odds ratio (95% CI), p-			
	Widowed	208	21.2	0.33 (0.24-0.47)	<0.001	0.85 (0.54-1.32)	0.466
Employment (most recent)	Unemployed	4956	47.9	Reference		Reference	
	Employed	341	39.6	0.71 (0.57-0.89)	0.003	0.65 (0.51-0.83)	<0.001
	In education	311	39.6	0.71 (0.56-0.90)	0.004	0.78 (0.61-1.02)	0.065
	Retired	7	14.3	0.18 (0.02-1.51)	0.114	0.51 (0.06-4.65)	0.547
ADL impairment	Absent	4700	41.9	Reference		Reference	

Factor	Group	Number in	Prevalence of two or more	Association with two or more negative symptoms: odds ratio (95% CI), p-			
	Present	2283	46.3	1.20 (1.08-1.32)	<0.001	1.29 (1.13-1.47)	<0.001
Social impairment	Absent	4432	42.7	Reference		Reference	
	Present	2533	44.4	1.07 (0.97-1.18)	0.158	0.93 (0.82-1.05)	0.258
Delusions / hallucinations	Absent	3904	41.9	Reference		Reference	
	Present	3077	45.0	1.14 (1.03-1.25)	0.009	1.23 (1.10-1.38)	<0.001
Depression	Absent	4976	45.2	Reference		Reference	
	Present	2014	38.8	0.77 (0.69-0.85)	<0.001	0.69 (0.61-0.78)	<0.001

Factor	Group	Number in	Prevalence of two or more	Association with two or more negative symptoms: odds ratio (95% CI), p-			

Appendix 4 Table 3: Association between number of negative symptoms ascertained prior to 2011 and mental health hospital admission, re-admission and duration of admission in 2011 in patients aged under 40 years and patients aged over 40 years

	Inpatient admission (odds ratio, 95% CI)*	**Re-admission within 12 months of inpatient admission (odds ratio, 95% CI)***	**Duration of inpatient admission (days; B-coefficient, 95% CI)****
Associations with 2 or more negative symptoms (binary variable) in patients aged between 16 and 39 years.	n=3315	n=792	n=785
Unadjusted	1.36 (1.17-1.59)	1.82 (1.36-2.43)	25.4 (6.2, 44.6)
1. Age and sex	1.40 (1.20-1.63)	1.88 (1.40-2.54)	20.8 (1.5, 40.1)
2. Model 1 plus marital status and employment	1.25 (1.06-1.46)	1.70 (1.24-2.31)	15.0 (-4.9, 34.9)
3. Model 2 plus delusions / hallucinations, and depression	1.22 (1.03-1.43)	1.68 (1.23-2.29)	14.5 (-5.5, 34.5)
Associations with 2 or more negative symptoms (binary variable) in patients aged over 40 years.	n=4361	n=820	n=805
Unadjusted	1.41 (1.20-1.65)	1.61 (1.21-2.16)	22.1 (5.2, 39.1)

	Inpatient admission (odds ratio, 95% CI)*	**Re-admission within 12 months of inpatient admission (odds ratio, 95% CI)***	**Duration of inpatient admission (days; B-coefficient, 95% CI)****
1. Age and sex	1.33 (1.13-1.56)	1.56 (1.16-2.08)	26.7 (9.8, 43.6)
2. Model 1 plus marital status and employment	1.26 (1.07-1.49)	1.48 (1.10-1.99)	24.3 (7.2, 41.4)
3. Model 2 plus delusions / hallucinations, and depression	1.24 (1.05-1.45)	1.48 (1.10-1.99)	24.4 (7.5, 41.4)

*Logistic regression; **Linear regression

Age x negative symptoms (binary variable) interaction term $p>0.05$ for all model.

Appendix 4 Table 4: Percentage of patients admitted to hospital in 2011 by number of negative symptoms (n=7678)

Number of negative symptoms	**Number of patients**	**Percentage admitted to hospital in 2011 (%)**
0	3408	21.7
1	1121	18.9
2	974	22.7
3	717	27.2
4	492	28.1
5	382	32.5
6 or more	584	36.8

Appendix 4 Table 5: Percentage of patients readmitted to hospital following discharge in 2011 by number of negative symptoms (n=1612)

Number of negative symptoms	**Number of patients**	**Percentage admitted to hospital in 2011 (%)**
0	612	29.9
1	195	34.4
2	213	40.4
3	176	44.9
4	131	43.5
5	119	47.1
6 or more	166	44.6

Appendix 4 Table 6: Median duration of admission amongst mental health inpatients with schizophrenia in 2011 by number of negative symptoms (n=1,609)

Number of negative symptoms	**Number of patients**	**Median duration of admission in days (inter-quartile range)**
0	696	30.0 (13- 73)
1	200	37.5 (17-86)
2	194	46.0 (17- 87)
3	165	40.0 (21- 89)
4	116	48.0 (16- 107)
5	110	51.5 (20- 128)
6 or more	128	56.5 (53- 178)

Appendix 5- Additional table for Chapter 9

Appendix 5 Table 1- Pearson's correlation coefficients between symptoms domains

	Negative	Positive	Manic	Disorganisation	Catatonic	Depressive
Negative	1	0.67	0.62	0.63	0.51	0.65
Positive	--	1	0.69	0.73	0.48	0.59
Manic	--	--	1	0.66	0.41	0.67
Disorganisation	--	--	--	1	0.49	0.59
Catatonic	--	--	--	--	1	0.46
Depressive	--	--	--	--	--	1

Bibliography

Aagaard, J. et al., 2016. Clinically useful predictors for premature mortality among psychiatric patients visiting a psychiatric emergency room. *International Journal of Social Psychiatry*, 62(5), pp.462–470.

Adam, D., 2013. Mental health: On the spectrum. *Nature*, 496(7446), pp.416–418.

Ajnakina, O. et al., 2018. Utilising symptom dimensions with diagnostic categories improves prediction of time to first remission in first-episode psychosis. *Schizophrenia Research*, 193, pp.391–398.

Allardyce, J. et al., 2007. Do symptom dimensions or categorical diagnoses best discriminate between known risk factors for psychosis? *Social Psychiatry and Psychiatric Epidemiology*, 42(6), pp.429–437.

Alloy, L.B. et al., 2005. The psychosocial context of bipolar disorder: Environmental, cognitive, and developmental risk factors. *Clinical Psychology Review*, 25(8), pp.1043–1075.

American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* 5th Editio., Washington, DC: American Psychiatric Association.

Anon, 1983. *Scale for the Assessment of Negative Symptoms*, Iowa City: University of Iowa Press.

Antolík, J., 2005. Automatic annotation of medical records. *Studies in health technology and informatics*, 116, pp.817–822.

Arango, C., Garibaldi, G. & Marder, S.R., 2013. Pharmacological approaches to treating negative symptoms: A review of clinical trials. *Schizophrenia Research*, 150(2–3), pp.346–52.

Avidan B, Sonnenberg A, Giblovich H, S.S., 2001. Reflux symptoms are associated with psychiatric disease. *Alimentary pharmacology & therapeutics*, 15(12), pp.1907–12.

Axelrod, B.N., Goldman, R.S. & Alphs, L.D., 1993. Validation of the 16-item Negative Symptom Assessment. *Journal of psychiatric research*, 27(3), pp.253–258.

Bagney, A. et al., 2013. Negative Symptoms and Executive Function in Schizophrenia: Does Their Relationship Change with Illness Duration? *Psychopathology*, 46(4), pp.241–248.

Bahorik AL, Satre DD, Kline-Simon AH, Weisner CM, C.C., 2017. Serious mental illness and medical comorbidities: findings from an integrated health care system. *Jorunal of Psychosomatic Research*, 100, pp.35–45.

Barnett, K. et al., 2012. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *The Lancet*, 380(9836), pp.37–43.

Bell MD, Lysaker PH, Beam-Goulet JL, Milstein RM, L.J., 1994. Five- component model of schizophrenia: assessing the factorial invariance of the Positive and Negative Syndrome Scale. *Psychiatry Res.*, 52(3), p.295–303.

Bellmaker R, 2004. Bipolar Disorder. *New England Journal of Medicine*, 351(5), pp.476–486.

Bill, F. et al., 2017. Cardiovascular disease and diabetes in people with severe mental illness : causes , consequences and pragmatic. *Punto Farmacológico*, 115(3), pp.1–11.

Blanchard JJ, Kring AM, H.W. et al, 2011. Toward the next generation of negative symptom assessments: the collaboration to advance negative symptom assessment in schizophrenia. *Schizophr Bull*, 37, p.291–9. doi:10.1093/schbul/sbq104.

Bleuler, M., 1963. Conception of Schizophrenia Within the Last Fifty Years and Today [Abridged]. *Proceedings of the Royal Society of Medicine.*, 56(10), pp.945–952.

Bobes, J. et al., 2010. Prevalence of negative symptoms in outpatients with schizophrenia spectrum disorders treated with antipsychotics in routine clinical practice: findings from the CLAMORS study. *The Journal of clinical psychiatry*, 71(3), pp.280–6.

Bowersox, N.W. et al., 2012. Cause-specific mortality among Veterans with serious mental illness lost to follow-up. *General Hospital Psychiatry*, 34(6), pp.651–653.

Brown, C. et al., 2010. Predictors of mortality in patients with serious mental illness and co-occurring type 2 diabetes. *Psychiatry Research*, 177(1–2), pp.250–254.

Brown S, M.C., 2012. Predictors of death from natural causes in schizophrenia: 10-year follow-up of a community cohort. *Social psychiatry and psychiatric epidemiology*, 47(6), pp.843–7.

Callaghan, R.C. et al., 2014. Patterns of tobacco-related mortality among individuals diagnosed with schizophrenia, bipolar disorder, or depression. *Journal of Psychiatric Research*, 48(1), pp.102–110.

Capasso, R.M. et al., 2008. Mortality in schizophrenia and schizoaffective disorder : An Olmsted County , Minnesota cohort : 1950 – 2005. , 98, pp.287–294.

Carney, C.P. & Jones, L.E., 2006. The influence of type and severity of mental illness on receipt of screening mammography. *Journal of General Internal Medicine*, 21(10), pp.1097–1104.

Castagnini, A., Foldager, L. & Bertelsen, A., 2013. Excess mortality of acute and transient psychotic disorders: Comparison with bipolar affective disorder and schizophrenia. *Acta Psychiatrica Scandinavica*, 128(5), pp.370–375.

Chang C-K, Hayes RD, Broadbent M, Hotopf M, Davies E, Møller H, S.R., 2014. A Cohort Study on Mental Disorders, Stage of Cancer at Diagnosis and Subsequent Survival. *BMJ Open*, 4, p.e004295. doi: 10.1136/bmjopen-2013-004295.

Chang, C.-K. et al., 2010. All-cause mortality among people with serious mental illness (SMI), substance use disorders, and depressive disorders in southeast London: a cohort study. *BMC psychiatry*, 10, p.77.

Chang CK, Hayes RD, Perera G, Broadbent MT, Fernandes AC, Lee WE, Hotopf M, S.R., 2011. Life Expectancy at Birth for People with Serious Mental Illness and Other Major Disorders from a Secondary Mental Health Care Case Register in London. *PLoS ONE*, 6(5), p.e19590.

Chang, W.C. et al., 2011. Persistent negative symptoms in first-episode schizophrenia: a prospective three-year follow-up study. *Schizophrenia research*, 133(1–3), pp.22–28.

Chapman, W.W. et al., 2011. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. *Journal of the American Medical Informatics Association*, 18(5), pp.540–543.

Chesney, E, Goodwin, GM, Fazel, S., 2014. Risks of all-cause and suicide mortality in mental disorders : a meta-review. *World Psychiatry*, 13, pp.153–160.

Chiang CL, 1984. *The Life Table and Its Applications*, Malabar, FL: Robert E. Krieger Publishing Company.

Chmielewski, M. et al., 2014. Openness to Experience, Intellect, Schizotypal Personality Disorder, and Psychoticism: Resolving the Controversy. *Journal of personality disorders*.

Cohen, C.I. et al., 2013. Prevalence of Negative Symptoms and Associated Factors in Older Adults With Schizophrenia Spectrum Disorder. *The American Journal of Geriatric Psychiatry*, 21(2), pp.100–107.

Cooke BK, Magas LT, Virgo KS, Feinberg B, A.A. and J.F., 2007. Appendectomy for appendicitis in patients with schizophrenia. *Am J Surg*, 193, p.41–48.

Cooper S, Crawford M, Etherington A, Jayakumar S, Lemmey S, Marsh R, Maynard E, S.D. and Z.K., 2012. Report of the National Audit of Schizophrenia (NAS) 2012. *Royal College of Psychiatrists.*

Corrigan, P.W., Druss, B.G. & Perlick, D.A., 2014. The Impact of Mental Illness Stigma on Seeking and Participating in Mental Health Care. *Psychological Science in the Public Interest*, 15(2), pp.37–70.

Cowen, P., 2018. *Shorter Oxford Textbook of Psychiatry*, Oxford University Press, Incorporated.

Crawford, M.J. et al., 2012. Group art therapy as an adjunctive treatment for people with schizophrenia: multicentre pragmatic randomised trial. *Bmj*, 344(feb28 4), pp.e846–e846.

Cross-Disorder Group of the Psychiatric Genomics Consortium., 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, 381(9875), pp.1371–9.

Crump, C. et al., 2013. Comorbidities and mortality in bipolar disorder: a Swedish national cohort study. *JAMA psychiatry*, 70(9), pp.931–9.

Cuesta, M.J. & Peralta, V., 2001. Integrating psychopathological dimensions in functional psychoses: a hierarchical approach. *Schizophrenia Research*, 52(3), pp.215–229.

Cunningham H, Tablan V, Roberts A, et al., 2013. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS Comput Biol*, 9, p.e1002854.

Das-Munshi, J. et al., 2017. Ethnicity and excess mortality in severe mental illness: a cohort study. *The Lancet Psychiatry*, 4(5), pp.389–399.

Das-Munshi, J. et al., 2016. Type 2 diabetes mellitus in people with severe mental illness: inequalities by ethnicity and age. Cross-sectional analysis of 588 408 records from the UK. *Diabetic medicine : a journal of the British Diabetic Association*.

Davidson S, Judd F, Jolley D, Hocking B, Thompson S, H.B., 2001. Cardiovascular risk factors for people with mental illness. *Aust N Z J Psychiatry*, 35, pp.196–202.

Davies SC, 2013. *Annual Report of the Chief Medical Officer 2013: Public mental health priorities: Investing in the evidence.*, Department of Health.

Davis, W.A. et al., 2015. The interactive effects of type 2 diabetes mellitus and schizophrenia on all-cause mortality: The Fremantle Diabetes Study. *Journal of Diabetes and its Complications*, 29(8), pp.1320–1322.

Demjaha, A. et al., 2009. Combining dimensional and categorical representation of psychosis: the way forward for DSM-V and ICD-11? *Psychological Medicine*, 39(12), p.1943.

Department of Health., 2013. National survey of Investment in Mental Health Services. , pp.1596–11. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/140098/FinMap2012-NatReportAdult-0308212.pdf, last accessed- 16th January 2018.

Der-Avakian A & Markou A, 2012. The Neurobiology of Anhedonia and Other Reward-Related Deficits. *Trends in Neurosciences*, 35.(1), pp.68–77.

Dickerson, F. et al., 2014. Mortality in Schizophrenia : Clinical and Serological Predictors. *Schizophr Bull*, 40(4), pp.796–803.

Dominguez, M.-G. et al., 2010. Early expression of negative/disorganized symptoms predicting psychotic experiences and subsequent clinical psychosis: a 10-year study. *The American journal of psychiatry*, 167(9), pp.1075–82.

Dorning, H. Davies, A. Blunt, I.., 2015. Focus on: People with mental ill health and hospital use. *The health Foundation- Nuffield Trust*. Available at: http://www.qualitywatch.org.uk/sites/files/qualitywatch/field/field_document/QualityWatch_Mental_ill_health_and_hospital_use_full_report.pdf, last accessed- 16th January 2018.

Druss BG, Zhao L, Von ES, Morrato EH, M.S., 2011. Understanding excess mortality in persons with mental illness: 17-year follow up of a nationally representative US survey. *Med Care*, 49, pp.599–604.

Dutta, R. et al., 2007. Suicide and other causes of mortality in bipolar disorder: a longitudinal study. *Psychological medicine*, 37(6), pp.839–47.

Embi, P.J. et al., 2005. Development of an electronic health record-based Clinical Trial Alert system to enhance recruitment at the point of care. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, pp.231–235.

Ernst, R. et al., 2013. Increasing mortality gap for patients diagnosed with schizophrenia over the last three decades — A Danish nationwide study from 1980 to 2010. *Schizophrenia Research*, 146, pp.22–27.

Fanton JH, Mcintyre RS, C.L., 2011. Psychiatric issues for patients with renal disease. *http://www.psychiatrictimes.com/articles/psychiatric-issues-patients-renal-disease, last accessed- 16th January 2018.*

Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang CK, Jackson RG, Roberts A, Tsang J, Soncul M, Liebscher J, Stewart R, C.F., 2013. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. *BMC medical informatics and decision making*, 13(1), p.71.

Fiedorowicz, J.G. et al., 2009. Manic/Hypomanic Symptom Burden and Cardiovascular Mortality in Bipolar Disorder. *Psychosomatic Medicine*, 71(6), pp.598–606.

Fiedorowicz, J.G. et al., 2014. Vascular Mortality in Participants of a Bipolar Genomics Study. *Psychosomatics*, 55(5), pp.485–490.

Fors, B.M. et al., 2016. Mortality among persons with schizophrenia in Mortality among persons with schizophrenia in Sweden : An epidemiological study. *Nordic Journal of Psychiatry*, 61(4), pp.252–259.

Foussias, G., Agid, O. & Remington, G., 2011. *Handbook of Schizophrenia Spectrum Disorders, Volume II* M. Ritsner, ed., Dordrecht: Springer Netherlands.

Frayne SM, Halanych JH, Miller DR, Wang F, Lin H, Pogach L, et al, 2014. Disparities in Diabetes Care. , 165, pp.2631–2638.

Fusar-Poli P, Papanastasiou E, S.D. et al, 2014. Treatments of negative symptoms in schizophrenia: meta-analysis of 168 randomized placebo-controlled trials. *Schizophr Bull*, 41, p.892–9. doi:10.1093/schbul/sbu170.

Gaebel, W. & Zielasek, J., 2015. Focus on psychosis. *Dialogues in Clinical Neuroscience*, 17(1), pp.9–18.

Gigerenzer G & Gaissmaier W, 2011. Heuristic decision making. *Annu Rev Psychol.*, 62, pp.451–82.

Gilbert, E.A. et al., 2000. Concurrent validity of negative symptom assessments in treatment refractory schizophrenia: relationship between interview-based ratings and inpatient ward observations. *Journal of psychiatric research*, 34(6), pp.443–447.

Glick, H.A., Li, P. & Harvey, P.D., 2015. The relationship between Positive and Negative Syndrome Scale (PANSS) schizophrenia severity scores and risk for hospitalization: an analysis of the CATIE Schizophrenia Trial. *Schizophrenia research*, 166(1–3), pp.110–4.

Goodwin, G.M. et al., 2016. Evidence-based guidelines for treating bipolar disorder: revised third edition Recommendations from the British Association for Psychopharmacology. *Journal of psychopharmacology (Oxford, England)*, 30(6), pp.495–553.

Gorrell G, Jackson R, Roberts A, et al., 2013. Finding negative symptoms of schizophrenia in patient records. In *Proceedings of NLP Med Biol Work (NLPMedBio), Recent Adv Nat Lang Process (RANLP), Hissar, Bulg*. pp. 9–17.

Hardoon, S. et al., 2013. Recording of Severe Mental Illness in United Kingdom Primary Care, 2000–2010 K. Laws, ed. *PLoS ONE*, 8(12), p.e82365.

Hayes, R.D. et al., 2012. Associations between symptoms and all-cause mortality in individuals with serious mental illness. *Journal of Psychosomatic Research*, 72(2), pp.114–119.

Hayes, R.D. et al., 2015. The effect of clozapine on premature mortality: An assessment of clinical monitoring and other potential confounders. *Schizophrenia Bulletin*, 41(3), pp.644–655.

Hayes RD, Chang CK, Fernandes AC, Begum A, To D, Broadbent M, Hotopf M, S.R., 2012. Functional status and all-cause mortality in serious mental illness. *PloS one*, 7(9), p.e44613.

Health and Social Care Information Centre., 2013. Mental Health Bulletin: Annual report from MHMDS returns – England 2011–12, initial national figures. https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-bulletin/mental-health-bulletin-annual-report-from-mhmds-returns-england-2011-12-further-analysis-and-organisation-level-data , [Accessed March 2017].

Health and Social Care Information Centre, 2013. HES-MHMDS Data Linkage Report: Summary Statistics, 2011/12 (Experimental Statistics). , 12(June 2013), pp.1–24.

Heckers, S., 2011. Bleuler and the neurobiology of schizophrenia. *Schizophrenia Bulletin*, 37(6), pp.1131–1135.

Hersh, W.R., 2007. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. *The American journal of managed care*, 13(6 Part 1), pp.277–278.

DE Hert M, Correll CU, Bobes J, Cetkovich-Bakmas M, Cohen D, Asai I, Detraux J, Gautam S, Möller HJ, Ndetei DM, Newcomer JW, Uwakwe R, L.S., 2011. Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 10(1), pp.52–77.

De Hert, M. et al., 2011. Physical illness in patients with severe mental disorders. II. Barriers to care, monitoring and treatment guidelines, plus recommendations at the system and individual level. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 10(2), pp.138–151.

De Hert M et al, 2009. Cardiovascular disease and diabetes in people with severe mental illness position statement from the European Psychiatric Association (EPA). *European Psychiatry*, 24, pp.412–424.

Hjorthøj, C. et al., 2015. Association between alcohol and substance use disorders and all-cause and cause-specific mortality in schizophrenia, bipolar disorder, and unipolar depression: A nationwide, prospective, register-based study. *The Lancet Psychiatry*, 2(9), pp.801–808.

Hoang, U., Goldacre, M.J. & Stewart, R., 2013. Avoidable mortality in people with schizophrenia or bipolar disorder in England. *Acta psychiatrica Scandinavica*, 127(3), pp.195–201.

Hoang, U., Stewart, R. & Goldacre, M.J., 2011. Mortality after hospital discharge for people with schizophrenia or bipolar disorder: retrospective study of linked English hospital episode statistics, 1999-2006. *BMJ (Clinical research ed.)*, 343, p.d5422.

Howes, O., Mccutcheon, R. & Stone, J., 2016. Europe PMC Funders Group Glutamate and dopamine in schizophrenia : an update for the 21 st century. J Psychopharmacol , 29(2), pp.97–115.

Howes, O.D. et al., 2017. The Role of Genes, Stress, and Dopamine in the Development of Schizophrenia. *Biological Psychiatry*, 81(1), pp.9–20.

Howes OD & Kapur S, 2009. The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr Bull.*, 35(3), pp.549–62.

Høye, A., Jacobsen, B.K. & Hansen, V., 2011. Increasing mortality in schizophrenia: Are women at particular risk? A follow-up of 1111 patients admitted during 1980-2006 in Northern Norway. *Schizophrenia Research*, 132(2–3), pp.228–232.

Hunter, R. & Barry, S., 2012. Negative symptoms and psychosocial functioning in schizophrenia: neglected but important targets for treatment. *European psychiatry : the Journal of the Association of European Psychiatrists*, 27(6), pp.432–436.

Insel, T. et al., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, 167(7), pp.748–751.

Jackson, R. et al., 2013. Finding Negative Symptoms of Schizophrenia in Patient Records. Proceedings of the Workshop on NLP for Medicine and Biology associated with RANLP 2013, pages 9–17

Jackson, R.G. et al., 2017. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open*, 7(1), pp.1–10.

Jacobs R, Gutacker N, Mason A, Goddard M, Gravelle H, Kendrick T, G.S., 2015. Determinants of hospital length of stay for people with serious mental illness in England and implications for payment systems: A regression analysis. *BMC Health Services Research*, 15(1), pp.1–16.

Jacobs, R., Gutacker, N. & Mason, A., 2015. Do higher primary care practice performance scores predict lower rates of emergency admissions for persons with serious mental illness? : an analysis of secondary panel data. *Health Services and Delivery Research*, 3(16).

Jacobs R, B.E.I. of crisis resolution and home treatment teams on psychiatric admissions in E.B.J.P. 2011;199:71–6., 2011. Impact of crisis resolution and home treatment teams on psychiatric admissions in England. *Br J Psychiatry*, 199, p.71–6.

Jager, M. et al., 2009. Prediction of symptom remission in schizophrenia during inpatient treatment. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry*, 10(4 Pt 2), pp.426–434.

Jaspers, K., 1973. *Allgemeine Psychopathologie.*, Berlin, Germany: Springer.

Jones DR, Macias C, Barreira PJ, Fisher WH, Hargreaves WA, H.C., 2004. Prevalence, Severity, and Co-Occurrence of Chronic Physical Health Problems of Persons With Serious Mental Illness. *Psychiatric services*, 55(11), pp.1250–1257.

Joukamaa M, 2001. Mental disorders and cause-specific mortality. *The British Journal of Psychiatry*, 179(6), pp.498–502.

Kay, S.R., Fiszbein, a & Opler, L. a, 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin*, 13(2), pp.261–76.

Kendrick T, 1996. Cardiovascular and respiratory risk factors and symptoms among general practice patients with long-term mental illness. *Br J Psychiatry 1996;169:733-9*, 169, pp.733–9.

Kessing LV, Vradi E, McIntyre RS, A.P., 2015. Causes of decreased life expectancy over the life span in bipolar disorder. *Journal of Affective Disorders*, 180, pp.142–7.

Ketter, T.A. & D., C.K., Bipolar and Related Disorders. *The American Psychiatric Publishing Textbook of Psychiatry, Sixth Edition.*

Ketter TA, 2010. Handbook of Diagnosis and Treatment of Bipolar Disorder. In Washington, DC, American Psychiatric Publishing.

Kim, W. et al., 2017. Mortality in Schizophrenia and Other Psychoses : Data from the South Korea National Health Insurance Cohort , 2002 – 2013. , (11), pp.835–842.

Kimhy, D. et al., 2006. The Factorial Structure of the Schedule for the Deficit Syndrome in Schizophrenia. *Schizophrenia Bulletin* , 32(2), pp.274–278.

Kirkpatrick B, Strauss GP, N.L. et al, 2011. The brief negative symptom scale: psychometric properties. *Schizophr Bull*, 37, p.300–5. doi:10.1093/schbul/sbq059.

Kirkpatrick, B. et al., 2006. The NIMH-MATRICS Consensus Statement on Negative Symptoms. *Schizophrenia Bulletin* , 32(2), pp.214–219.

Kisely et al., 2013. Reducing all-cause mortality among patients with psychiatric disorders: a population-based study. *CMAJ*, 185(1), pp.E50-6.

Kisely S, Ehrlich C, Kendall E, L.D., 2015. Using avoidable admissions to measure quality of care for cardiometabolic and other physical comorbidities of psychiatric disorders: A population-based, record-linkage analysis. *Canadian Journal of Psychiatry*, 60(11), pp.497–506.

Kisely S, Quek LH, Pais J, Lalloo R, Johnson NW, L.D., 2011. Advanced dental disease in people with severe mental illness : systematic review and meta-analysis. , pp.187–193.

Kisely S, C.L. and W.Y., 2009. Treatment of ischaemic heart disease and stroke in individuals with psychosis under universal healthcare. *Br J Psychiat*, 195, p.545–550.

Kraeplin, 1893. *Dementia Praecox and Paraphrenia*, Chicago Medical Book Co., Chicago.

Kredentser, M.S. et al., 2014. Cause and rate of death in people with schizophrenia across the lifespan: A population-based study in Manitoba, Canada. *Journal of Clinical Psychiatry*, 75(2), pp.154–160.

Kring, A.M., 2013. The Clinical Assessment Interview for Negative Symptoms (CAINS): Final Development and Validation. *American Journal of Psychiatry*, 170(2), p.165.

Lahti, M. et al., 2012. Cardiovascular morbidity, mortality and pharmacotherapy in patients with schizophrenia. *Psychological medicine*, 42(11), pp.2275–85.

Lamkin J, Nguyen PT, Coverdale JH, G.M., 2016. Towards a Definition of “Self-Neglect” in Psychiatric Patients: Descriptions of a Case Series. *Psychiatric Quarterly*, (3), pp.1–8.

Laursen, T.M. et al., 2014. Cardiovascular drug use and mortality in patients with schizophrenia or bipolar disorder: a Danish population-based study. *Psychological medicine*, 44(8), pp.1625–37.

Laursen, T.M., Munk-Olsen, T. & Gasse, C., 2011. Chronic somatic comorbidity and excess mortality due to natural causes in persons with schizophrenia or bipolar affective disorder. *PLoS ONE*, 6(9).

Laursen TM, Munk-Olsen T, Agerbo E, G.C. and M.P., 2009. Somatic hospital contacts, vasive cardiac procedures, and mortality from heart disease in patients with severe mental disorder. *Archs Gen Psychiat*, 66, p.713–720.

Laursen TM, Munk-Olsen T, Nordentoft M, M.P., 2007. Increased mortality among patients admitted with major psychiatric disorders: a register-based study comparing mortality in unipolar depressive disorder, bipolar affective disorder, schizoaffective disorder, and schizophrenia. *Journal of Clinical Psychiatry*, 68, pp.899–907.

Lawrence D, Hancock KJ, K.S., 2013. The gap in life expectancy from preventable physical illness in psychiatric patients in Western Australia: retrospective analysis of population based registers. *BMJ (Clinical research ed.)*, 346(May), p.f2539.

Lawrence D, Holman CDJ, J.A. and H.M., 2003. Death rate from ischaemic heart disease in Western Australian psychiatric patients 1980–1998. *Br J Psychiat*, 182, p.31–36.

Lawrence D, Jablensky AV, Holman CDJ, P.T., 2000. Mortality in Western Australian psychiatric patients. *Soc Psychiatry Psychiatr Epidemiol*, 35, pp.341–7.

Lawrence D, Mitrou F, Z.S., 2009. Smoking and mental illness: results from population surveys in Australia and the United States. *BMC Public Health*, 9, p.285.

Lawrence, D. & Kisely, S., 2010. Inequalities in healthcare provision for people with severe mental illness. *Journal of psychopharmacology (Oxford, England)*, 24(4 Suppl), pp.61–8.

De Leon J, D.F., 2005. A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophr Res*, 76((2-3)), pp.135–57.

Leucht, S. et al., 2007. Physical illness and schizophrenia: A review of the literature. *Acta Psychiatrica Scandinavica*, 116(5), pp.317–333.

Leverich, G.S. & Post, R.M., 2006. Course of bipolar illness after history of childhood trauma. *The Lancet*, 367(9516), pp.1040–1042.

Li, Y., Bontcheva, K. & Cunningham, H., 2009. Adapting SVM for data sparseness and imbalance: a case study in information extraction. *Natural Language Engineering*, 15(02), pp.241–271.

Lichtenstein, P. et al., 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*, 373(9659), pp.234–239.

Lopez-Morinigo JD, Fernandes AC, Chang CK, Hayes RD, Broadbent M, Stewart R, David AS, D.R., 2014. Suicide completion in secondary mental healthcare: a comparison study between schizophrenia spectrum disorders and all other diagnoses. *BMC Psychiatry*, 14, p.213.

Lukasiewicz, M. et al., 2013. Young Mania Rating Scale: how to interpret the numbers? Determination of a severity threshold and of the minimal clinically significant difference in the EMBLEM cohort: YMRS severity cutoff. *International Journal of Methods in Psychiatric Research*, 22(1), pp.46–58.

Matevosyan, N., 2010. Oral health of adults with serious mental illnesses: A review. *Community Mental Health Journal*, 46(6), pp.553–562.

McGurk, S.R. et al., 2000. Relationship of cognitive functioning, adaptive life skills, and negative symptom severity in poor-outcome geriatric schizophrenia patients. *The Journal of neuropsychiatry and clinical neurosciences*, 12(2), pp.257–264.

McIntyre RS et al., 2007. mMedical comorbidity in bipolar disorder: re-prioritizing unmet needs. *Curr Opin Psychiatry.*, 20(4), p.406–16.

McPherson S, Barbosa-Leiker C, Daratha K, Short R, McDonell M, G, Alicic R, Roll J, T.K., 2014. Association of Co-Occurring Serious Mental Illness with Emergency Hospitalization in People with Chronic Kidney Disease. *American Journal of Nephrology*, 39, pp.260–267.

Medici, C. et al, 2015. Mortality and secular trend in the incidence of bipolar disorder. *Journal of affective disorders*, 183, pp.39–44.

Millard, C. & Wessely, S., 2014. Parity of esteem between mental and physical health: Means different things to different people, making it difficult to enforce. *BMJ (Online)*, 349(November), pp.1–2.

Miller BJ, Paschall CB 3rd, S.D., 2006. Mortality and medical comorbidity among patients with serious mental illness. *Psychiatric Services*, 57(10), pp.1482–1487.

Mitchell AJ, Delaffon V, L.O., 2012. Let's get physical: improving the medical care of people with severe mental illness. *Advances in Psychiatric Treatment*, 18(3), pp.216–225.

Mitchell AJ, M.D. and D.C., 2009. Quality of medical care for people with and without comorbid mental illness and substance misuse: systematic review of comparative studies. *Br J Psychiat*, 194, p.491–499.

Moller, H.J. et al., 2000. Long-term course of schizophrenic, affective and schizoaffective psychosis: focus on negative symptoms and their impact on global indicators of outcome. *Acta psychiatrica Scandinavica. Supplementum*, (407), pp.54–57.

Morden, N.E. et al., 2012. Eight-year trends of cardiometabolic morbidity and mortality in patients with

schizophrenia. *General Hospital Psychiatry*, 34(4), pp.368–379.

Morgan, M.G. et al., 2003. Prospective analysis of premature mortality in schizophrenia in relation to health service engagement : a 7 . 5-year study within an epidemiologically complete , homogeneous population in rural Ireland. *Psychiatry Research*, 117(03), pp.127–135.

Muntaner C, Eaton WW, Miech R, O.P., 2004. Socioeconomic position and major mental disorders. *Epidemiol Rev*, 26, pp.53–62.

Nasrallah, H. et al., 2006. Low rates of treatment for hypertension, dyslipidemia and diabetes in schizophrenia: data from the CATIE schizophrenia trial sample at baseline. *Schizophr Res*, 86, pp.15–22.

National Confidential Enquiry into Patient Outcome and Death (NCEPOD), 2017. Treat as One. Bridging the gap between mental and physical healthcare in general hospitals, p.110. http://www.ncepod.org.uk/2017mhgh.html, [Accessed May 3, 2018].

Naylor C et al, 2012. Long-term Conditions and Mental Health: The cost of co-morbidities. *The King's Fund and Centre for Mental Health.* Available at: https://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/long-term-conditions-mental-health-cost-comorbidities-naylor-feb12.pdf [Accessed April 1, 2017].

NHS Choices, 2017. Treatment of Schizophrenia. Available at: http://www.nhs.uk/Conditions/Schizophrenia/Pages/Treatment.aspx [Accessed April 15, 2017].

NHS England, 2015. Mental Health Taskforce. Available at: www.england.nhs.uk/mentalhealth/taskforce/ [Accessed January 5, 2017].

NHS England, Reducing mortality for people with serious mental illness (SMI). *2015*. Available at: www.england.nhs.uk/ourwork/forward-view/sop/red-prem-mort/smi/ . [Accessed January 15, 2017].

NICE, 2009. Clinical guidance 178- Psychosis and schizophrenia in adults: prevention and management. Available at: https://www.nice.org.uk/guidance/cg178 [Accessed March 6, 2017].

NICE, 2014. Psychosis and schizophrenia in adults: prevention and management. Available at: https://www.nice.org.uk/guidance/cG178/ifp/chapter/Assessment [Accessed January 18, 2017].

Nikiforou, A., Ponirou, P. & Diomidous, M., 2013. Medical data analysis and coding using natural language processing techniques in order to derive structured data information. *Studies in health technology and informatics*, 190, pp.53–55.

Nordentoft M, Wahlbeck K, Hällgren J, Westman J, Osby U, Alinaghizadeh H, Gissler M, L.T., 2013. Excess mortality, causes of death and life expectancy in 270,770 patients with recent onset of mental disorders in Denmark, Finland and Sweden. *PLoS ONE*, 8, p.e55176.

Office for National Statistics, Causes of Death. Available at: http://www.ons.gov.uk/ons/taxonomy/search/index.html?nscl=Causes+of+Death&nscl-orig=Causes+of+Death&content-type=Dataset&content-type=Reference+table&sortDirection=DESCENDING&sortBy=pubdate [Accessed August 1, 2015a].

Office for National Statistics, Period- and cohort-life expectancy tables. Available at: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/lifetables/period-and-cohort-life-expectancy-tables/2010-based/new-component.xls [Accessed August 1, 2016b].

Office for National Statistics, 2013. Statistical bulletin: Deaths Registered in England and Wales: 2012. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregistrationsummarytables/2013-07-10#summary [Accessed June 1, 2014].

Olfson M, Ascher-Svanum H, F.D. et al., 2011. Predicting psychiatric hospital admission among adults with schizophrenia. *Psychiatr Serv*, 62, p.1138–45.

Olfson, M. et al., 2015. Premature Mortality Among Adults With Schizophrenia in the United States. *JAMA psychiatry*, 72(12), pp.1172–81.

Van Os, J. et al., 1997. Psychopathological syndromes and familial morbid risk of psychosis. *The British Journal of Psychiatry*, 170(3), pp.241–246.

Van Os, J. & Kapur, S., 2009. Schizophrenia. *Lancet*, 374(9590), pp.635–645.

van Os, J., Kenis, G. & Rutten, B.P.F., 2010. The environment and schizophrenia. *Nature*, 468, p.203.

Osborn DPJ, Nazareth I, K.M., 2007. Physical activity, dietary habits and Coronary Heart Disease risk factor knowledge amongst people with severe mental illness: a cross sectional comparative study in primary care. *Soc Psychiatry Psychiatr Epidemiol*, 42, pp.787–93.

Ösby, U. et al., 2016. Mortality trends in cardiovascular causes in schizophrenia, bipolar and unipolar mood disorder in Sweden 1987-2010. *European Journal of Public Health*, 26(5), pp.867–871.

Palmer BA, Pankratz VS, B.J., 2005. The lifetime risk of suicide in schizophrenia: a reexamination. *Arch Gen Psychiatry*, 62, pp.247–53.

Pan, Y.-J. et al., 2017. Transformation of excess mortality in people with schizophrenia and bipolar disorder in Taiwan. *Psychological Medicine*, (May), pp.1–11.

Pan, Y. et al., 2016. Hospital treatment , mortality and healthcare costs in relation to socioeconomic status among people with bipolar affective disorder.BJPsych Open,2(1), pp.10–17.

Pan, Y.J. et al., 2016. Three-year mortality in relation to early hospitalization and number of outpatient clinic visits in people with newly diagnosed bipolar disorder. *General Hospital Psychiatry*, 43(71), pp.32–37.

Partti K et al, 2015. Lung function and respiratory diseases in people with psychosis: population-based study. *British Journal of Psychiatry*, 207, pp.37–45.

Patel R, Jayatilleke N, J.R. et al., 2014. Investigation of negative symptoms in schizophrenia with a machine learning text-mining approach. *Lancet*, 383, p.S16.

Patel R, Lloyd T, J.R. et al, 2015. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. *BMJ Open*, 5, p.e007504 doi:10.1136/bmjopen-2014-007504.

Patel, R. et al., 2015. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. *BMJ Open*, 5(9), p.e007619.

Pendlebury, J. & Holt, R.I., 2010. Managing Diabetes in People with Severe Mental Illness. *Journal of Diabetes Nursing*, 14(9), pp.328–338.

Perera G, Broadbent M, Callard F, Chang CK, Downs J, Dutta R, Fernandes A, Hayes RD, Henderson M, Jackson R, Jewell A, Kadra G, Little R, Pritchard M, Shetty H, Tulloch A, S.R., 2016. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: Current status and recent enhancement of an Electronic Mental Health Record-derived data resource. *BMJ Open*, 6(3).

Phelan JC, Link BG, T.P., 2010. Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. *J Health Soc Behav*, 51, pp.S28-40.

Picchioni, M. & Murray, R., 2007. Schizophrenia. *BMJ*, 335(7610), pp.91–95.

Potuzak, M. et al., 2012. Categorical vs dimensional classifications of psychotic disorders. *Comprehensive Psychiatry*, 53(8), pp.1118–1129.

Ramsey, C.M. et al., 2013. Lifetime manic spectrum episodes and all-cause mortality: 26-year follow-up of the NIMH epidemiologic catchment area study. *Journal of Affective Disorders*, 151(1), pp.337–342.

Rathore, S.S. et al., 2008. Mental disorders, quality of care, and outcomes among older patients hospitalized with heart failure: an analysis of the national heart failure project. *Archives of general psychiatry*, 65(12), pp.1402–8.

Reininghaus, U. et al., 2015. Mortality in Schizophrenia and Other Psychoses : A 10-Year Follow-up of the Æ SOP First-Episode Cohort. *Schizophrenia Bulletin*, 41(3), pp.664–673.

Russo, M. et al., 2014. Association between symptom dimensions and categorical diagnoses of psychosis: A cross-sectional and longitudinal investigation. *Schizophrenia Bulletin*, 40(1), pp.111–119.

Saha S, C.D. and M.J., 2007. A systematic review of mortality in schizophrenia Is the differential mortality gap worsening over time? *Archs Gen Psychiat*, 64, pp.1123–1131.

Schizophrenia Working Group of the Psychiatric Genomics Consortium., 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), pp.421–7.

Schneider, 1959. *Clinical Psychopathology*, Grune and Stratton, New York.

Schoepf D, Uppal H, Potluri R, H.R., 2014. Physical comorbidity and its relevance on mortality in schizophrenia: A naturalistic 12-year follow-up in general hospital admissions. *European Archives of Psychiatry and Clinical Neuroscience*, 264(1), pp.3–28.

Scott, D. & Happell, B., 2011. The high prevalence of poor physical health and unhealthy lifestyle behaviours in individuals with severe mental illness. *Issues in mental health nursing*, 32(9), pp.589–97.

Sebat J, Levy DL & McCarthy SE, 2009. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends in genetics*, 25(12), pp.528–35.

Sendt, K.-V., Tracy, D.K. & Bhattacharyya, S., 2015. A systematic review of factors influencing adherence to antipsychotic medication in schizophrenia-spectrum disorders. *Psychiatry Research*, 225(1), pp.14–30.

Stewart R, Soremekun M, Perera G, Broadbent M, Callard F, Denis M, Hotopf M, Thornicroft G, L.S., 2009. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. *BMC Psychiatry*, 9(1), p.51.

Stuart H, 2003. Violence and mental illness: an overview. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 2(2), pp.121–4.

Talaslahti, T. et al., 2012. Mortality and causes of death in older patients with schizophrenia. *International journal of geriatric psychiatry*, 27(11), pp.1131–7.

Talaslahti, T. et al., 2015. Patients with very-late-onset schizoprhenia-like psychosis have higher mortality rates than elderly patients with earlier onset schizophrenia. *International Journal of Geriatric Psychiatry*, 30(5), pp.453–459.

Taylor, C.L. et al., 2015. The characteristics and health needs of pregnant women with schizophrenia compared with bipolar disorder and affective psychoses. *BMC psychiatry*, 15, p.88.

Tenback. et al., 2012. All-cause mortality and medication risk factors in schizophrenia: a prospective cohort study. *Journal of clinical psychopharmacology*, 32(1), pp.31–5.

Thornicroft G, 2013. Premature death among people with mental illness. *BMJ*, 346, p.f2969.

Tiihonen, J. et al., 2016. Mortality and cumulative exposure to antipsychotics, antidepressants, and benzodiazepines in patients with schizophrenia: An observational follow-up study. *American Journal of Psychiatry*, 173(6), pp.600–606.

Toffol, E. et al., 2015. Lithium is associated with decrease in all-cause and suicide mortality in high-risk bipolar patients: A nationwide registry-based prospective cohort study. *Journal of Affective Disorders*, 183, pp.159–165.

Tondo, L. et al., 2016. Comparison of psychotic bipolar disorder, schizoaffective disorder, and schizophrenia: An international, multisite study. *Acta Psychiatrica Scandinavica*, 133(1), pp.34–43.

Tzeng NS, Hsu YH, Ho SY, Kuo YC, Lee HC, Yin YJ, Chen HA, Chen WL, Chu WC, H.H., 2015. Is schizophrenia associated with an increased risk of chronic kidney disease ? A nationwide matched-cohort study. *BMJ Open*, (June 2009), pp.1–10.

Ucok, A., Serbest, S. & Kandemir, P.E., 2011. Remission after first-episode schizophrenia: results of a long-term follow-up. *Psychiatry research*, 189(1), pp.33–37.

Wallwork, R.S. et al., 2012. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. *Schizophrenia Research*, 137(1–3), pp.246–250.

Westman, J. et al., 2013. Cardiovascular mortality in bipolar disorder: a population-based cohort study in Sweden. *BMJ open*, 3(4), p.e002373.

White J, Gutacker N, Jacobs R, M.A., 2014. Hospital admissions for severe mental illness in England: Changes in equity of utilisation at the small area level between 2006 and 2010. *Social Science and Medicine*, 120, pp.243–251.

WHO, 2017. Schizophrenia. Available at: http://www.who.int/mental_health/management/schizophrenia/en/ [Accessed May 5, 2017].

Wing JK, Beevor AS, Curtis RH, et al., 1998. Health of the Nation Outcome Scales (HoNOS). Research and development. *Br J Psychiatry*, 172, pp.11–18.

Wing JK, Beevor AS, C.R. et al, 1998. Health of the Nation Outcome Scales (HoNOS). Research and development. *Br J Psychiatry*, 172, pp.11–18.

Woodhead, C. et al., 2016. Cervical and breast cancer screening uptake among women with serious mental illness: a data linkage study. *BMC cancer*, 16(1), p.819.

World Health Organization, 2010. The ICD-10 Classification of Mental and Behavioural Disorders. *http://www.who.int/classifications/icd/en/bluebook.pdf, last accessed 2nd February 2017*.

Wu, C.-S. et al., 2015. Complications and mortality in patients with schizophrenia and diabetes: population-based cohort study. *The British journal of psychiatry : the journal of mental science*, 207(5), pp.450–7.

Wu S-I et al, 2013. Association between schizophrenia and bipolar disorder with diagnostic procedures, revascularization, and inpatient mortality following acute myocardial infarction. *Psychosom Med*, 75, pp.52–9.

Yi-Hua C, H.-C.L. and H.-C.L., 2011. Poor Clinical Outcomes Among Pneumonia Patients With Schizophrenia. , 37(5), pp.1088–1094.

Yogeswary K, 2014. Schizoaffective Disorder: An overview,. *International Journal of Clinical Psychiatry*, 2(1), pp.11–15.

Yong, A. & Pearce, S., 2013. A Beginner ' s Guide to Factor Analysis : Focusing on Exploratory Factor Analysis. , 9(2), pp.79–94.

Zai, G. et al., 2017. A review of molecular genetic studies of neurocognitive deficits in schizophrenia. *Neuroscience and Biobehavioral Reviews*, 72, pp.50–67.

Zolnierek CD, 2009. Non-psychiatric hospitalization of people with mental illness: systematic review. *Journal of advanced nursing*, 65(8), pp.1570–83.